Vol. 59, No. 6 November, 1952

Psychological Review

EDITED BY
CARROLL C. PRATT
PRINCETON UNIVERSITY

CONTENTS

Walter V. Bingham 1880–1952.....HAROLD E. BURTT 403

Introduction to a Theory of Motivation in Learning.....JOHN P. SEWARD 405

Language, Psychology and Linguistics.....DAVID L. OLMSTED & OMAR KHAYYAM MOORE 414

An Analysis of Meaning.....CLYDE E. NOBLE 421

The Role of Observing Responses in Discrimination Learning. Part I.....L. BENJAMIN WYCKOFF, JR. 431

The Measurement of Discriminative Behavior.....LEE S. CHRISTIE 443

The Brain Analogy: A Discussion.....H. EDGAR COBURN 453

A "Communication Analysis" of Concept Learning.....CARL I. HOVLAND 461

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

The *Psychological Review* is devoted primarily to articles in the field of general and theoretical psychology. This area is obviously difficult to define and delimit, but in view of the large number of manuscripts sent to the editor on all kinds of topics an attempt has to be made to draw the line somewhere.

Ordinarily manuscripts that run to more than about 7500 words are not accepted. This policy is followed partly in an effort to reduce lag of publication and partly from the conviction that brevity which is not inconsistent with clarity is the best way to present an argument.

If an author is prepared to pay for the cost of publishing his article, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors.

Tables, footnotes, and references as well as text of manuscripts should be typed double-spaced throughout.

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.
$5.50 volume $1.00 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

H. P. Bingham.

VOL. 59, NO. 6 NOVEMBER, 1952

THE PSYCHOLOGICAL REVIEW

WALTER V. BINGHAM
1880–1952

In the death of Walter Bingham applied psychology has lost one of its leading figures. Dr. Bingham began his career in the era when psychology was mostly concerned with fundamentals, but he was among the earliest to see the practical possibilities of the subject. This insight, coupled with his skill in administrative and promotional work and in getting groups to work together, enabled him to launch a number of ventures with far-reaching consequences for the whole of psychology. While I was a student at Dartmouth I first knew him when he came there as Assistant Professor in 1910. He had secured his Ph.D. at Chicago two years earlier with prior study at Beloit, Berlin and Harvard. He established a small laboratory along Titchenerian lines, but within a few years he was at the Carnegie Institute of Technology organizing the Bureau of Salesmanship Research. He assembled a notable group for that pioneer effort in cooperative research. The impact of such an experience on the subsequent careers of the young psychologists working with him is hard to estimate. At the American Psychological Association meeting at Pennsylvania State College two years ago when some of his students came together with later associates to do him honor, it was certainly impressive just to look around the group.

Dr. Bingham left Carnegie Institute in 1924 and for ten years put a major portion of his effort into the Personnel Research Federation, of which he was Director. Toward the end of this period he took up a part-time connection with Stevens Institute of Technology where he was concerned with vocational guidance and personnel psychology. He continued there till the clouds began to gather for World War II, when he went to Washington for the rest of his career.

Some of his most notable contributions were in the field of military psychology. In 1917 there was not much applied psychology and no military psychology at all. One of the first steps was to set up the Committee on Classification of Personnel in the Army of which Bingham was made executive secretary. Presently he was commissioned Lieutenant Colonel in the Personnel Board of the General Staff. That was a high rank for those days. Only one psychologist ranked as a full Colonel in the first war. When World War II broke out Bingham decided to remain in a civilian capacity, thinking that he could thus do a better job. He was chairman of another committee on Classification of Military Personnel throughout the war period. At the same time he was chief psychologist in the Adjutant General's office. To those of us who had occasion to visit him there during the war it was obvious that he was in a key position and that he had far-reaching influence in the program. He was awarded the Secretary of War's Emblem for Exceptional Civilian Service. After the war he continued in a consulting capacity to the Army General Staff and to the Secretary of Defense.

In connection with professional organizations he carried his share of the load:

secretary of the American Psychological Association, 1911–1914; president of the American Association of Applied Psychologists, 1941; president of the New York State Association for Applied Psychology, 1938; first chairman of the Division of Anthropology and Psychology of the National Research Council, 1919; and on the Board of the Psychological Corporation for many years.

For some years he edited the *Journal of Personnel Research,* was consulting editor to several other journals, and contributed various journal articles. His *Aptitudes and Aptitude Testing,* perhaps his major work, was being revised at the time of his death. In collaboration he wrote *Procedures in Employment Psychology* and *How to Interview.* The last thing he published, his "colyum" in *Personnel Psychology,* described his experiences last summer at the two International Conferences in Sweden. Appropriately enough the article ends on a note of minimizing the difference between applied and general psychology.

It is always difficult at a time like this to find just the right word for the more personal aspects of a man's life. He was friendly and pleasant in social contacts. He had a contagious enthusiasm. It was hard to keep him down. He was active in professional meetings in recent years in spite of considerable physical discomfort. In the postscript to his last article mentioned above he was cheerfully recommending an oxygen tent to his friends.

It may well be that even more important than his notable publications are the intangible contributions which he has made in giving counsel, encouragement and stimulation to younger men working in his field of interest. It is in this capacity that many of us will remember him most vividly.

HAROLD E. BURTT
The Ohio State University

INTRODUCTION TO A THEORY OF MOTIVATION IN LEARNING

BY JOHN P. SEWARD

University of California, Los Angeles

I. Problem

In 1942 (21) I reported an attempt to test Guthrie's finality theory of reinforcement. In that experiment rats that were removed from the apparatus as soon as they pressed a bar made the response more frequently than controls. But rats that were first given a food pellet on pressing the bar responded at a higher level than those that were simply removed. This result did not fit easily into Guthrie's theory, especially since bar pressing was more nearly the final response in the box for the removed group than for the fed-and-removed group. Thorndike's or Hull's version of reinforcement seemed adequate to the numerical data but somehow failed to do justice to the observed behavior of the animals. A naive observer would have said that the fed rats *wanted to eat* more than the removed rats *wanted to get out of the box.* The problem remained: how did the animals' wants determine their behavior?

The following year (22) I tried to answer the question by means of a choice point with food on one side and a blind alley on the other. Briefly, I argued that after a number of trials the animal responded to the two paths with expectations, or *surrogates,* of food and dead-end, respectively; that since he was hungry the food surrogate was already aroused; that self-stimulation was therefore stronger on the side toward food; and that strength of stimulus determined choice. This explanation raised two questions: (1) How was the food surrogate attached to hunger in the first place? (2) Aside from the two surrogates could we be sure that *total* stimulation was stronger in the direction of food? Might not the frustration generated by the dead-end surrogate produce more "stimulus-intensity dynamism" (10) than would result from food anticipation?

Guthrie would have a ready answer to the first question; as to the second he might hold that conditioned withdrawal responses would prevent entrance into the blind. But take the analogous case of a rat in a T with shock through the floor everywhere except in the goal-box. Suppose a non-correction method with removal from either endbox after 30 seconds. Here there is no obvious goal surrogate to add stimulus intensity; rather the expectation that shock will continue is more excitatory than that it will cease. Moreover, responses to continued shock in the wrong endbox probably resemble responses at the choice point more closely than do those in the box without shock; the former should therefore be less likely to inhibit forward locomotion.[1]

In 1947 (23), with Hull's principles (9) as a guide, I proposed what was essentially an elaboration of the earlier theory in an attempt to account for the facts of latent learning. In particular I applied the theory to Blodgett's (1) experiment in which he introduced a food reward in a rectangular maze with one long and one short path to the goal-box. The sudden drop in errors was attributed to an abrupt increase of one factor in the food drive, a component

[1] Guthrie could still resort to finality through change of situation to explain learning under these conditions. But we have seen reason to doubt that this principle is sufficient by itself.

aroused by food itself, conditioned to the goalbox, and aroused as a food surrogate at the choice point. Since this factor was held to be stronger on the side nearer the goal, reaction potential and the probability of choice to that side were also greater.

But although more dressed up, the theory was no less vulnerable than before. Again we ask: What if, on the critical trial, instead of putting food in the goalbox Blodgett had shocked his rats throughout the rest of the maze? Since the drive surrogate would now be more strongly aroused by long-path cues we should have to predict fewer choices of the short one. The prediction is worth checking, but it is probably wrong.

My next attempt to improve the theory (25) took the form of the following postulate: *Reaction potential between S and R* ($_sE_R$) *is increased by the simultaneous activation of the surrogate response to a satisfier* (rs_G). Certainly this proposition covered a wider range of conditions, but only through being too vague to be useful. Without a clear definition of a satisfier we are unable to use the postulate for prediction. For example, does rs_G increase $_sE_R$ in the absence of a relevant need, i.e., unless the organism is already set for the satisfier? A more precise statement of limiting conditions was called for.

Such a statement was attempted in a recent discussion (26) of Hull's revised postulate set (10). Here rs_G, as a goal set, became a secondary motive differentially conditioned to choice point cues. Choice was determined when these cues *facilitated* the goal set. This process I called *tertiary motivation* and proposed it as a substitute for secondary reinforcement with its habit-strengthening connotation.

Does the concept of tertiary motivation meet the two-fold requirement of breadth and precision? Unfortunately not. It came from analysis of the goal-seeking aspect of behavior; application to avoidance is awkward at best. Consider the accompanying diagram, which attempts to fit a typical appetite (food-hunger) and a typical aversion (pain-escape) into the proposed scheme of motivation. For the appetite we can say that promotion of D_2 "reinforces" a response, but not for the aversion. And if "reinforcement" depends on enhancement of a goal set, then the reinforcement of an avoidant response calls for a D_4!

Motivation	Food-hunger	Pain-escape
Primary (D_1)	Hunger stimulus	Pain stimulus
Secondary (D_2)	Set to eat	Fear
Tertiary (D_3)	Facilitated set	Set to escape

The gist of the problem can be stated as follows: In order to account for the discontinuities of behavior brought about by manipulating rewards I have made reinforcement affect performance rather than learning. But whichever a reinforcer does, it *is* essentially a satisfier, a need reducer. How, then, can the anticipated reduction of a drive cause the *increase* of $_sE_R$ necessary to response selection? [2] In positive cathexes involving a characteristic consummatory response the increase of strength can be attributed to its surrogate. But in negative ones, dominated by what the animal seeks to escape rather than attain, how does an organism avoid catastrophe?

Two courses of action are possible: either to adopt one principle for appetites and another, such as Guthrie's, for aversions; or to re-examine the nature of motivation in hope of finding a common mechanism. In this paper we choose the second course.

STRUCTURE OF MOTIVES

Motivation refers to those processes most essential to the initiation, mainte-

[2] Hull resolved this dilemma by having drive reduction increase the habit-strength component of $_sE_R$. But this solution runs into difficulties, as elsewhere shown (26).

nance, and direction of activity. Of what use is a concept so broadly defined? This one is needed to account for the segmental character of behavior, the property that supplied Muenzinger (19) with a psychological unit of description. An organism, as he pointed out, typically moves from a disturbing situation to one of relative quiescence. What Muenzinger calls a "start-to-end phase" we shall call a *motivated act* (*M*).

A *motive* (*m*), the inner core of *M*, is harder to identify. Not directly observable, it has the status of a hypothetical construct in that it refers to a central neural process. It is equivalent to Morgan's (17) concept of the *central motive state*, to which he assigned the properties of perpetuating itself, evoking general activity, and "priming" and emitting specific reactions. Through observed intercorrelations among instigators, satisfiers, and intervening events specific *m*s can be identified and classified. Our purpose is not to classify them but to describe their common structure.

If central processes are to be handled by an *S–R* psychology they must be identified by the *S*s arousing them or the *R*s they lead to, or both. A convenient device for this purpose is the *surrogate response* (*rs*), elsewhere defined "on a molecular level . . . as a neural process occurring in the central connections between *S* and *R*; on a molar level, as the functional equivalent of the whole *S–R* circuit" (24, p. 282). Since the essentials of the learning process take place in the brain and possibly in the spinal cord, it is reasonable to suppose that internuncial elements achieve some degree of functional autonomy; i.e., they can perform associative functions with greatly lessened external support. Thus the central counterpart of a response (perceptual, locomotor, emotional, or other) may be aroused in the absence of its original stimulus and in turn provide stimulation for associated responses. Such a construct is especially useful in dealing with acquired motives and in deriving the "purposiveness" (29) of motivated acts. It does not, of course, rule out the probability that *rs* is often supplemented by response-produced cues.

We are now ready for a general description of *m*. Let us start with a few typical examples. A hungry boy feels the pangs of an empty stomach and is the first to hear the chime of a distant ice cream truck. His lonely sister is more likely to hear the mail man or the telephone. A hiker caught in a downpour thinks of a warm fire and dry clothes; on a hot summer day he yearns for a cool dip. If we can generalize from such cases it appears that most motives can be analyzed into two components, one associated with annoyers, the other with satisfiers. I suggest that we refer to these parts as *moments* of *m* and that we call the first moment *drive* and the second *goal*. They will shortly be defined as surrogate responses.[3]

All *m*s do not lend themselves readily to description in these terms. Some seem to be all drive: some physical discomforts, the boredom of a solitary child, the paralyzing fear of a hunted animal. Others seem to consist entirely of goal: the desire to play, to climb a mountain, to see a movie, to eat candy after dinner. Such examples fall readily into the twofold classification we have already borrowed from Craig via Tolman (29), *viz.*, into appetites and aversions. An appetite may be defined as an *m* instigated, in part, by the absence of a certain

[3] What justification is there for endowing *rs*s with properties of *m*? A deduction to this effect must rest on prior assumptions as to the nature of ${}_{s}E_{R}$ and the conditions under which it becomes capable of prolonging itself and of priming and setting off other *R*s. Since such speculations are fruitless at present we shall simply make the motivational status of the *rs* moments a first-order assumption and seek support in whatever verifiable deductions may follow.

source of stimulation and ended by responding to it in a certain way. An aversion is an *m* instigated by the presence of a certain stimulus source and ended by its absence. To decide whether aversions and appetites differ in basic structure or merely in emphasis we shall have to examine the part played by learning in their development. But first let me introduce a few more symbols with brief definitions of the concepts they stand for.

S_I—*instigator:* a stimulus pattern regularly followed by the onset of *M*.

S_D—*primary drive stimulus:* an S_I producing or produced by a homeostatic imbalance, e.g., a painful stimulus, empty stomach contractions, a dry throat.

rs_D—*surrogate drive stimulus; drive,* for short: an *rs* initially aroused by S_D; the first moment constituting *m*.

Srs_D—*secondary drive stimulus:* a stimulus conditioned to rs_D, therefore able to arouse it in the absence of S_D; e.g., a dentist's drill, Culler's buzzer.

G—*goal state:* an inferred state of balance in the organism after *M* is over.

S_G—*primary goal stimulus:* a collective term for the inner and outer *S*s entering into *G*.

R_G—*primary goal response:* a collective term for the overt and covert *R*s taking place in *G*.

rs_G—*surrogate goal response:* an *rs* initially aroused as part of the neural circuit between S_G and R_G.

R_g—*subgoal response:* any *R* closely and consistently followed by *G*; e.g., eating by a hungry organism.

rs_g—*surrogate subgoal response.*

rs_G—$rs_G + rs_{g_1} + \cdots + rs_{g_n}$—*surrogate goal hierarchy; goal,* for short: a central neural complex made up of goal and subgoal surrogates; the second moment constituting *m*.

Srs_G—*secondary goal stimulus:* a stimulus conditioned to rs_G, thus capable of arousing it before S_g or *G* occurs; e.g., the sight of food, Pavlov's tuning fork.

The problem of how motives are acquired may be dealt with under two main heads: (1) acquiring of goals; (2) acquiring of drives.

1. *Acquiring of goals.* A goal, as defined above, is acquired through the conditioning of rs_G to new aspects of the situation. These include both S_D and other *S*s. By virtue of conditioning to rs_D, rs_G is shifted forward to the beginning of *M* and thus conditioned to still earlier *S*s. In the same way rs_g, such as the *rs* for eating or pressing a bar or entering a white box, is acquired. A well developed goal-moment thus consists not only of rs_G but of a hierarchy of rs_gs. It is for this reason that we need the inclusive symbol rs_G.

Hull gave the classical account of this process in his 1931 paper (7), using the constructs of the fractional goal reaction and its proprioceptive cue. The relative prominence of rs_G, and of certain subgoals as compared with others, remains to be explained. Hull (8) supplemented his description by the concept of the goal gradient, according to which *R*s are conditioned more strongly to *S*s the closer they are to what we call *G* and what he called a reinforcing state of affairs. To me it seems likely that the factors of finality and frequency are more fundamental than reinforcement. Since *G* is always the final state superseding *m*, its surrogate rs_G is in the ideal position for learning. By the same principles, supplemented by a postulate to be introduced below, the more closely and consistently an R_g is followed by *G* the more frequently it will be made and the more strongly rs_g will be conditioned.[4]

[4] Our account parts company with that of Murphy (20), who calls goal-learning *canalization* and sets it apart from conditioning, and with that of Tolman (30), who distinguishes among *positive cathexes, positive equivalence beliefs,* and *field expectancies.* Since this is not the place for a critical discussion of viewpoints the present treatment is necessarily dogmatic.

Will the same description hold equally for appetites and aversions? One apparent distinction between them is that in appetites more than aversions one R_g is relatively fixed by nature for all members of a species and primed by some S_D or chemical agent. Although it is questionable whether the R_gs of eating, drinking, or copulating are more reflex than is withdrawing from a noxious stimulus, the generalization probably holds. But what of it? Does it mean that R_gs culminating in a consummatory reflex will be more easily or stably learned than those without such a biological anchor? Actually learned avoidant responses are notably persistent (2, 16) and most cases of abnormal fixation involve fear.[5] The chief significance of an innate R_g is that the organism's early trial-and-error behavior will be more limited in the event of certain cyclic disturbances than of adventitious ones,—an arrangement with obvious biological advantages. The important thing for us is that in both appetites and aversions the focusing of wants on specific satisfiers must be learned and that the same learning principles seem to apply.

2. *Acquiring of drives.* This process is envisaged as the conditioning of rs_D, originally aroused by S_D, to other Ss; a simple contiguity principle seems adequate to account for it. Here we face the question: Can both appetitive and aversive drives be conditioned? Common sense says yes. One road sign warns me of sudden death, another "makes me thirsty." But we need more convincing evidence.

So far it has been found much easier to condition fear in the laboratory than to condition hunger or thirst. One trouble with appetites is that to test for rs_D in the absence of S_D it is necessary to satiate the animal, thus producing a most unfavorable condition for testing. Several studies, however, have shown that secondary reinforcers based on hunger or thirst help animals maintain old habits or learn new ones even when satiated (3, 13, 28). If we assume that secondary reinforcers are effective only if they reduce some drive, these results imply that a conditioned drive is operating. Increased hoarding by rats as a result of *earlier* frustration (11, 12) suggests a similar interpretation.

Even if we grant that an appetite can function without S_D we are still not sure whether it is instigated by Srs_D or Srs_G. This doubt was settled for shock-avoidance by Brown and Jacobs' (4) repetition of Miller's (15) classical study. But in an appetite the two moments are harder to isolate because of the early, if not innate, connection between S_D and R_g. Until some method is devised it will probably make for theoretical simplicity to assume that the conditioned S_I, in appetites as in aversions, is Srs_D.[6] On this assumption the difference between them is reduced to one of degree.

In summary, we may think of a motive as a hypothetical construct made up of two moments, a surrogate drive stimulus (rs_D) and a surrogate goal hierarchy (rs_G) representing goal and available subgoal responses. These moments are integrated largely through learning, which consists of the conditioning of rs_G to rs_D and of each one to other Ss. Both appetites and aversions are held to conform to this description.

FUNCTION OF MOTIVES

With a generalized description of a motive at hand, we are ready to make

[5] It is interesting that what Maier (14) calls "behavior without a goal" may be used as a striking example of the tenacity with which an animal clings to a specific goal when caught in a "tight spot."

[6] Mowrer (18) takes a similar position with his concept of "drive-fear." The implications of Srs_G for a theory of functional autonomy of motives are intriguing but must be reserved for treatment elsewhere.

a start toward answering the question: how does it direct behavior? Since m is an intervening construct the question breaks down into two parts: (1) How do conditions affecting the organism change the functional characteristics of m? (2) How do these changes determine the organism's selection of responses?

In answer to part 1, the strength of each moment can presumably be raised by unconditioned and conditioned stimuli and lowered by their removal. We must also recognize that the two moments are so closely related that a change in one involves a change in the other. Thus, since rs_G is conditioned to rs_D, its strength will vary as some positive function of drive. On the other hand, in some cases Srs_D may displace Srs_G, thus reducing goal intensity. In the same way independent variation of rs_G may react negatively on rs_D, since R_g and G are associated with the cessation of drive. These secondary relations will not be considered in this preliminary treatment.

As to part 2, there are a number of salient facts about behavior that seem to reflect the influence of motivation. One is that as the strength of drive increases the organism's responses become more vigorous. This effect can be partially handled for some drives by the general energizing action of Hull's *primary drive* (D) considered as a chemical agent. But since certain Rs are apparently more sensitized than others a more specific mechanism is needed. For this purpose it may help to recall that m, as a central impulse pattern, functions both as R and as S. In the course of trial-and-error behavior both moments become conditioned to successive Ss on the one hand and Rs on the other. To Hull's postulate VI regarding *stimulus-intensity dynamism* (V) we may add a tentative assumption as to the V of a motive (V_m). This assumption is that V_m is a linear function of the sum of intensities of rs_G and rs_D, i.e.,

$$V_m = a(rs_G + rs_D) + k. \qquad (1)$$

It follows that increments in m will step up the ${}_sE_R$ to any R conditioned to it.

It is equally clear, however, that not all Rs conditioned to m are equally likely to be made, but that those leading most directly to G have a decided edge. At first glance it may seem as if equation (1) will handle this fact, too. Since Ss closest to G are conditioned most strongly to rs_G their Rs should command the largest V_m. A second glance shows that V_m does not necessarily follow a goal gradient. For while rs_G is raised by goal-pointing Ss, rs_D may be equally raised by remote ones, so that both sets of Ss have the same effect on V_m. This difficulty is especially clear in the T-maze shock-escape problem already considered.

To predict goal-direction a further assumption is needed. We do not have far to seek. In view of the close integration of rs_D and rs_G, it would be surprising if they acted in complete independence. We therefore assume that V_m is a function not merely of their sum but also of their interaction. As a first approximation I suggest that it be made secondarily a positive linear function of the algebraic difference between rs_G and rs_D. Equation (1) takes the following revised form:

$$V_m = a(rs_G + rs_D) + b(rs_G - rs_D) + k. \qquad (2)$$

Algebraically this revision amounts to weighting rs_G more heavily than rs_D in determining V_m.

Equation (2) seems to contain the essential ingredients of an adequate substitute for the gradient of reinforcement without the embarrassments encountered by that construct. By means of the con-

cept of V_m as here defined we can handle the T-maze situation with ease, and with the aid of quantitative refinements we should be able to deduce the gradients of amount and delay of reward. Since G affects ${}_sE_R$ by way of V rather than ${}_sH_R$ it should also be possible to explain sudden changes in behavior with the introduction, increase, or decrease of reward (27).

One finding not included in the present theory is the "contrast effect" reported by Crespi (5) and Zeaman (31). These investigators found that when the amount of reward to which rats were trained was suddenly increased, the rats started sooner or ran faster than others trained to the larger amount from the beginning. Crespi also obtained a downward contrast effect by reducing the reward, but for Zeaman it was not significant. Another example of downward contrast was observed by Elliott (6) in the maze performance of rats when he changed their reward from bran mash to sunflower seed.

It is possible to derive this phenomenon by making one more extension of the formula for V_m—a refinement not too far-fetched in its own right. We may argue that if an increment in rs_G relative to rs_D raises V_m, the rate at which the increment occurs may have something to do with its effectiveness. The more rapid the increase the more V_m may be heightened. If this much be granted, then by the same token a sharp, uncompensated drop in rs_G should *lower* V_m more than a gradual one. On these assumptions we revise equation (2) as follows:

$$V_m = a(rs_G + rs_D) + b(rs_G - rs_D) + c[d(rs_G - rs_D)/dt] + \text{k}. \quad (3)$$

Applying this formula to Crespi's and Zeaman's results, we assume that, on the next trial after the augmented reward was found, starting box stimuli evoked an rs_G corresponding to the smaller amount superseded by an rs_G to the larger. A rapid increase in rs_G thus took place in the starting box. Theoretically the first two terms on the right-hand side of equation (3) brought the experimental animals to the level of the controls; the third term carried them above it. We also predict that as the earlier goal-trace became progressively weaker, the third term would approach zero and the E curve would drop back to the C level. So far no such study has been continued long enough to verify the prediction. Downward contrast is derived in the same way, though why it should be less striking than upward contrast is not explained. Incidentally, it should be possible to demonstrate both effects by shocking two groups in a straightaway and varying the degree of shock reduction in the endbox.

Further development, quantification, and testing of the theory must be left to the future. All that remains to be done now is to point out how the present formulation relates to the steps preceding it. The terms secondary and tertiary motivation have to be given new meanings for comparison with those indicated in the diagram above. Secondary motivation now refers to rs_D and rs_G in so far as they are activated by conditioned instigators; its strength is included in equation (1). Tertiary motivation was my attempt to single out the specific contribution to ${}_sE_R$ of the goal as the crucial factor in response selection. That concept is now embodied in the last two terms of equation (3). It is still essentially what it was originally intended to be: a substitute mechanism for secondary reinforcement.

SUMMARY

An attempt has been made to correct certain shortcomings noted in my previous efforts to solve the problem of rein-

forcement. A major concern was to develop a mechanism that would apply equally well to learning motivated by appetites and aversions.

The present "solution" analyzes all motives into two main parts, or *moments, a drive* (rs_D) and a *goal* (rs_G), instigated and integrated largely through conditioning. Their role in problem solving is envisaged as follows:

1. In determining the *stimulus-intensity dynamism* of a motive (V_m), rs_G is assigned greater weight than rs_D.

2. *S*s and *R*s near reward are conditioned more strongly to rs_G than those more remote.

3. Goal-pointing *S*s therefore generate the greater V_m and their *R*s command the greater ${}_sE_R$.

The theory is extended to provide for "contrast effects."

REFERENCES

1. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113–134.
2. BROGDEN, W. J. Retention of conditioned responses tested by experimental extinction. *Amer. J. Psychol.*, 1940, **53**, 285–288.
3. ——. Non-alimentary components in the food-reinforcement of conditioned forelimb-flexion in food-satiated dogs. *J. exp. Psychol.*, 1942, **30**, 326–335.
4. BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
5. CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.
6. ELLIOTT, M. H. The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, **4**, 19–30.
7. HULL, C. L. Goal attraction and directing ideas conceived as habit phenomena. PSYCHOL. REV., 1931, **38**, 487–506.
8. ——. The goal-gradient hypothesis and maze learning. PSYCHOL. REV., 1932, **39**, 25–43.
9. ——. *Principles of behavior*. New York: Appleton-Century, 1943.
10. ——. Behavior postulates and corollaries —1949. PSYCHOL. REV., 1950, **57**, 173–180.
11. HUNT, J. MCV., SCHLOSBERG, H., SOLOMON, R. L., & STELLAR, E. Studies of the effects of infantile experience on adult behavior in rats. I. Effects of infantile feeding frustration on adult hoarding. *J. comp. physiol. Psychol.*, 1947, **40**, 291–304.
12. MCCORD, F. The effect of frustration on hoarding in rats. *J. comp. Psychol.*, 1941, **32**, 531–541.
13. MACCORQUODALE, K., & MEEHL, P. E. "Cognitive" learning in the absence of competition of incentives. *J. comp. physiol. Psychol.*, 1949, **42**, 383–390.
14. MAIER, N. R. F. *Frustration: the study of behavior without a goal*. New York: McGraw-Hill, 1949.
15. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
16. ——. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 435–472.
17. MORGAN, C. T. *Physiological psychology*. New York: McGraw-Hill, 1943.
18. MOWRER, O. H. Comment on Estes' study: "Generalization of secondary reinforcement from the primary drive." *J. comp. physiol. Psychol.*, 1950, **43**, 148–151.
19. MUENZINGER, K. F. *Psychology: the science of behavior*. New York: Harper, 1942.
20. MURPHY, G. *Personality: a biosocial approach to origins and structure*. New York: Harper, 1947.
21. SEWARD, J. P. An experimental study of Guthrie's theory of reinforcement. *J. exp. Psychol.*, 1942, **30**, 247–256.
22. ——. Reinforcement in terms of association. PSYCHOL. REV., 1943, **50**, 187–202.
23. ——. A theoretical derivation of latent learning. PSYCHOL. REV., 1947, **54**, 83–98.
24. ——. The sign of a symbol: a reply to Professor Allport. PSYCHOL. REV., 1948, **55**, 277–296.
25. ——. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, **39**, 177–186.

26. ——. Secondary reinforcement as tertiary motivation: a revision of Hull's revision. PSYCHOL. REV., 1950, **57**, 362–374.
27. ——. Experimental evidence for the motivating function of reward. *Psychol. Bull.*, 1951, **48**, 130–149.
28. ——, & LEVY, N. Maze learning by satiated rats as a function of secondary reinforcement. *Amer. Psychol.*, 1951, **6**, 277–278.
29. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.
30. ——. There is more than one kind of learning. PSYCHOL. REV., 1949, **56**, 144–155.
31. ZEAMAN, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

[MS. received August 27, 1951]

LANGUAGE, PSYCHOLOGY AND LINGUISTICS

BY DAVID L. OLMSTED

Institute of Human Relations
Yale University

AND

OMAR KHAYYAM MOORE

Tufts College

In recent years psychologists[1] have come to be more and more interested in language and cue-producing responses in general. Adherents of each of the major learning-theories are attempting to work out rationales adequate, it is hoped, to account for the learning of the complex linguistic responses which come to form part of the response-repertoire of most human beings living in social groups. It has been established, of course, that language is learned, and thus is a legitimate concern of the learning-theorist and other psychologists. Psychologists' studies of language necessarily involve dealing with the same subject matter that is the concern of the linguist, whose interest has been increasingly in the development of a rigorous system for the description and formal analysis of language. To be sure, linguists have occasionally and uncritically adopted one or more of the current theories of acquisition, or concocted *ad hoc* explanations of language-acquisition to meet the requirement of the lectures on language-learning in an elementary class in linguistics. Some of these theories have unfortunately found their way into print. However, this tendency has shown a marked decrease since Bloomfield pointed out twenty years ago "that we can pursue the study of language without reference to any one psychological doctrine, and that to do so safeguards our results and makes them more significant to workers in related fields" (**2**, p. vii). In other words, the linguist is to devote himself to description and formal synchronic and diachronic analysis, irrespective of current theories as to how language is acquired. Insofar as the number of such *ad hoc* ventures has decreased, Bloomfield's remark has been salutary in effect. However, a too narrow interpretation of that remark can result in closing off areas of joint research. In view of the careful experimental work and notable achievements in theory construction on the part of psychologists and the recent advances in linguistic analysis, it may be that the time has come when the linguist and the psychologist can undertake some joint investigations with profit.

Just as linguists have occasionally created or uncritically adopted theories of language-acquisition, so also have psychologists, in the course of their investigations of language-learning and other matters, developed their own linguistic theories. Characteristic of these theories is the assumption of linguistic "units" like the "sound," "syllable," "word," and "sentence" as self-evident. They are nowhere defined in these theories.

[1] We wish to express our gratitude to Dr. Donald J. Lewis for his many helpful comments, from the psychologist's point of view, about the paper. Thanks are also due to Dr. Norman A. McQuown, and Dr. Samuel E. Martin, with whom the first-named author discussed some of the linguistic problems involved. Final responsibility for everything in the paper, however, rests with the authors.

As a matter of fact, linguists, after prolonged methodological difficulties, have come to the conclusion that these "folk-units" have no cross-cultural validity, that is to say, they can only be used by being redefined for each language described. For example, one would be hard put to it to discover comparable units in such diverse languages as Yoruba, Annamese and Yaruro if the folk-unit "word" is employed as the basis of description. Even if defined separately for each language, the units called "words" in each of these languages might well not be comparable, since the different linguistic criteria employed in the definition might result in a polymorphemic word for one language, whereas it might turn out that in a second language all words are monomorphemic. The implications of this lack of compatibility are obviously of great importance for the psychologist. For example, a study, based on the use of the notion "word," of the vocabulary-size of speakers of two languages might reveal great differences in word-count between them, whereas the actual number of learned discriminations might well be very nearly the same.

In order to point out the difficulties that arise when psychologists make up their own linguistics, we turn to an examination of George Miller's chapter, "Speech and Language," in the recently-published *Handbook of Experimental Psychology* (**12**). We singled out this work for special consideration precisely because it appears in a standard handbook that will be widely used, represents a summary of work accomplished and in progress rather than a narrowly individual contribution, and is at least as competently done as anything in the field with which we are acquainted.[2]

1. We find three kinds of difficulties with Miller's chapter: those that concern particulars, those of general theoretical import, and those having to do with bibliography.

1.0. Difficulties that concern particulars are, in turn, of three varieties: the factually false, the highly ambiguous, and the misleading.

1.1. *The factually false.* We take it as understood that empirical or contingent propositions, because they are not analytic in character, may be said to be true or false with only a certain degree of probability. We believe that there is ample evidence to indicate that the examples we cite are false. We shall present a number of such factually false propositions for purposes of illustration.

a. "One of these consonants, the glottal stop ʔ, is not used in normal English speech, but we should probably add two more: tš in choke, and dž in joke" (**12**, p. 795). Now we know of no dialect of English in which the glottal stop is *not* employed; the only reason it does not turn up in the frequency studies quoted by Miller is that these studies are based on the traditional orthography, and the glottal stop is not ordinarily written in any one consistent way in English. For example, some, perhaps most, American pronunciations of *button* [bʌ′ʔn], contain the glottal stop. The probability is high that every time George A. Miller says his name, "George *A*. Miller," with emphasis, he utilizes this by no means absent consonant at the beginning of what is

[2] We wish specifically to avoid the suggestion that our comments are singularly applicable to any particular person; we address them rather to all students of language problems. Since this paper was written, Dr. Miller's book, *Language and Communication*, has been published. It represents, in some respects, improvement over the chapter in the handbook, and, being an important work, deserves detailed consideration on its own merits, at a length outside the scope of this paper.

written with the initial *A*. in the conventional orthography.

In many dialects of American English, morphemes written with an initial vowel have two alternate phonetic shapes: One, occurring initially in an utterance, begins with a glottal stop; the other, in utterance-medial position, begins with the vowel.[3]

With respect to the initial phonemes of *choke* and *joke*, they are either sequences of /t/ plus /š/ and /d/ plus /ž/ as Miller indicates, in which case they are *not* to be added to the list of phonemes, or, they are unit phonemes /č/ and /ǯ/ (in which case they are *not* to be written with the symbols /tš/ and /dž/ which indicate consonant clusters, *not* unit phonemes). Such contrasts as *scorching vs. courtship* make it necessary to regard the latter alternative as the more probable. In short, Miller has either improperly analyzed or improperly employed the initial consonants of *choke* and *joke*.

b. "The classification of English vowel sounds is usually made in terms of the position of the tongue. In this way the back vowels of *boot* or *book* are distinguished from the front vowels in *beet* or *bit*" (**12**, p. 794). The classification of English vowels is made, as a matter of fact, in terms of tongue position and lip-rounding.[4] *Boot* contains not only a back vowel, but a rounded vowel, and not all back vowels are rounded, *cf.*, for example, the vowel of Western New York *fog*, which is low back and unrounded.

Considerations of space prevent us from giving additional instances of factually false propositions.

1.2. *The highly ambiguous.* (a) "The strongest evidence for preferred sequences comes from the casual observation that consonants and vowels *tend to* alternate in speech" (**12**, p. 797). It is difficult to determine whether this statement is trivial, almost a tautology; for, on the assumption that within any given idiolect there are consonants and vowels, variation in their occurrence is inevitable. If, on the other hand, as seems more likely, Miller intends to suggest that utterances are composed of sequences of vowels and consonants such that two or more vowels or consonants do not occur together, then this flies in the face of the empirical evidence. If Miller had been making his "casual observation" in Polish-speaking territory, he might have noted a reverse tendency, since medial consonant clusters of nine consonants occur in Polish. To be sure, there are languages, e.g., Yoruba, which have no consonant clusters, but English is not one of them. Since languages exhibit such diverse structures with regard to this matter, it seems difficult to determine what could possibly be meant by saying that vowels and consonants "tend to" approximate the CVCV form.

(b) "As the child matures he employs an increasingly wide range of sounds, and at 2.5 years the distribution is very similar to that for adults" (**12**, p. 799). Without dealing with the unfortunately arbitrary age suggested in the second part of the above proposition, an age for which adequate cross-cultural validation has not been collected, we turn to the expression "wide range of sounds." It is unclear, and crucially unclear, whether Miller uses "sound" to refer to a phoneme or to a noise produced by a given articulatory act. If the latter, this seems not to be in accord-

[3] We wish specifically to avoid the problem of the phonemic status of [ʔ]. For one treatment of the matter, and of a number of other linguistic complexities overlooked by Miller, see (**16**).

[4] Not to mention degree of tension, see (**16**).

ance with the findings of Leopold, (**11**) Velten, (**17**) Jakobson, (**10**) and Hockett, (**9**), who indicate that the babbling infant engages in a wider range of articulations than that of the adult, and that the number of different articulations decreases as the number of phonemic discriminations increases. In other words, the infant makes a great variety of noises that are not part of the language of his speech-community, and these, from the point of view of reinforcement theory, not being rewarded, are extinguished at the same time that the child gradually develops new contrasts where there were none, thus increasing the list of his *phonemes*. On the other hand, the articulations of the adult have relatively high habit strength, so that the wide range of articulations possible for him when he was a babbling infant are now found to be difficult when he attempts to learn a foreign language.

1.3. *The misleading.* (a) In a discussion pertaining to unit consonants, Miller states that certain ones "š, št, dž, ž are seldom used in any position" (**12**, p. 797). Although it is very likely true that these phonemes and clusters are rare in English, it is misleading to group them together in a discussion of unit phonemes, since it is unclear whether dž represents a cluster or a unit phoneme in Miller's scheme, and št is unambiguously a cluster. If the subject of clusters is to be broached at all, either directly or indirectly, then the distinction between unit phonemes and clusters should be made, since it is well known that unit phonemes and clusters are not usually similar in pattern.

(b) One of the most misleading features of the chapter has to do with the use Miller makes of the "word"-lists of Dewey, Thorndike, and others. Many years ago Bloomfield discussed such attempts to get at language through written sources: "If we take a large body of speech, we can count out the relative frequencies of phonemes and of combinations of phonemes. This task has been neglected by linguists and very imperfectly performed by amateurs, who confuse phonemes with printed letters" (**2**, p. 136–137). The difficulty with using so-called "word"-lists is that the orthography of English inaccurately and inconsistently reflects the phonemic pattern and, unless one works with a phonemic analysis constantly at hand, gross inaccuracies of the type listed in 1.1 are inevitable. More important, traditional "word"-lists, unless they are subjected to a morphological analysis, do not give anything like an accurate account of the basic linguistic signalling units, the morphemes. It is perhaps not entirely fortuitous that the morpheme concept, one of the central concepts of modern descriptive linguistics, is not to be found in Miller's chapter, or in most psychological studies of language. The somewhat extensive statistical operations performed with such dubious data as "word"-lists are consequently to be rated at something less than face value. Even Zipf, who is used as an authority by Miller, states, relative to Dewey's word count, the same one that is used by Miller: ". . . these cannot be said to represent any actually spoken language, nor was that the author's intent. . . . Let us repeat again, that the present law of frequency can only be exactly demonstrated in an *actually spoken dialect*" (**18**) (italics ours). Fortunately, presentday linguists have at last turned their attention to quantitative analysis of actually spoken dialects;[5] and it is therefore no longer necessary to rely on "word"—lists of

[5] *Cf.*, for example (**7**, **15**).

the kind utilized by Miller and others working in the same field.

2.0. *Difficulties of general theoretical import.* There are a number of distinctions that linguists, working with problems of the formal analysis of language in the last 100 years, have found it necessary to make. Long ago linguists discovered that the orthographies of those languages that are customarily written are usually imperfectly related to the sound system of what is spoken. Therefore there are certain things that can be said of an orthography that may not be said of the utterances it is supposed to represent. This then, is a fundamental and elementary distinction that all linguists, whether their concern be with historical problems or descriptive analysis, are agreed upon; it permeates the whole of linguistic science. Illustrative of the difficulties into which neglect of this distinction can lead one are those experienced by Newman in his article (**14**) dealing with the distribution of vowels and consonants in various languages. In working out the statistics, he finds it convenient to indicate consonants by 1 and vowels by 0, and indicates that English *the* would be represented by 110. Here is a classic example of the results of linguistic research being rendered invalid by confusion of letters in the orthography with the phonemes. For this utterance does not begin (phonemically) with /th/, but with /ð/, this fact being obscured by the English-speaker's custom of writing /ð/ with the two letters *th*, relying on knowledge of the language to distinguish between what is represented by *th* in *hothouse* and *other* (not to mention what it represents in *nothing*). The utterance *the* in Newman's scheme would thus be 10, rather than 110. Newman's method of following the orthography must have led to gross errors, if, as he indicates, he was interested in the *language* and not the *orthography*. For example, *through* would be in his system 1110011, while *to* would be 10, although phonemically they are, respectively, /θrúw/ and /túw/, and would thus be either 1100 and 100 or 1101 and 101, depending on the status of /w/ as vowel or consonant. Moreover, forms like *bough*, *cough*, *though*, *slough*, would be similarly represented, although different phonemically, while forms like *oh*, *sew*, *though*, *beau*, *hoe*, *owe*, *fro*, *faux* pas, although phonemically alike, would be differently represented.

It is incumbent upon anyone working in this general area to decide whether he wants to study orthographies or languages, since there are important differences between the two. If any given study be so crude that it makes no difference if the results for the one be taken for the results for the other (something that could only be ascertained by doing the study both orthographically and phonemically), it is the researcher's responsibility to point out the crudity of the results and to show further that the hypothesis that he was testing was likewise of sufficient crudity to justify using such gross methods.

For the last half century linguists have been preoccupied with arriving at descriptive units adequate for any spoken language. Through a process of trial and error they have found that certain units are unworkable, while other units that have been redefined and explicated have been rendered cross-culturally valid. Typical of these units is the phoneme, for which the definitive work by Bloch (**1**) is only the culmination of generations of work by others, and the morpheme, given its present state of explication by such workers as Bloomfield (**2**),

Harris (**4, 5, 6**) and Hockett (**8**). It seems to us extremely unfortunate for psychologists to recapitulate the weary phylogeny of linguistic development with all its blind alleys, and fruitless controversies. The difficulties of the psychologists are likely to be compounded by virtue of the fact that their linguistic naiveté is overlaid with the protective coloration of impressive statistical procedures.

To return to Miller's chapter, we are told that "verbal behavior proceeds word by word," and that "the choice of the word as the unit in this process (the study of language) is not crucial, for the same thing can be done with the 26 letters of the alphabet and a space." For the linguist this option is like that between Scylla and Charybdis, since the inapplicability of "letters" has been pointed out above, and the "word," not a unit having cross-cultural validity, is a folk-unit apparently defined as whatever is written in the traditional orthography with spaces around it.

There are a great many other important theoretical considerations which we would like to treat in detail. Among these is the concept of "verbal context," which Miller is at great pains to develop by means of "statistical approximations to English." This, as a first attempt, is creditable, but unnecessary, since the notion was clearly worked out by Harris (**5**) a decade ago. Other considerations of this sort are Newman's concept of "primitive languages" (one now discarded by anthropologists and linguists), Miller's assumption of the "syllable" as a valid linguistic unit, and his preoccupation with English, which leads him to fabricate generalizations which may well not prove to be the case when the evidence from other languages is in.

Matters concerning bibliography manifest something of a cultural lag of roughly twenty years with respect to matters linguistic. E.g., no important linguistic work and no work relating to the theory of descriptive linguistics whatsoever bearing a date later than 1930 is cited. As a matter of fact, most of the important linguistic work that appeared earlier than that date is omitted, e.g., the work of Sapir. Those psychologists[6] who are interested in pursuing the matter further may refer to the bibliography of works dealing with linguistics in the excellent work by the psychologist Carroll (**3**).

In summary we would suggest that inasmuch as psychologists have found it profitable to work with physiologists and inadvisable to invent their own physiology, so too we think there is much to be gained by a similar attitude toward linguistics, and we have prepared this paper in the interests of achieving that end.

REFERENCES

1. BLOCH, B. A set of postulates for phonemic analysis. *Language*, 1948, **24**, 3–46.
2. BLOOMFIELD, L. *Language*. New York: Henry Holt, 1933.
3. CARROLL, J. B. *A survey of linguistics and related disciplines*. (Multilithed.) Cambridge, Mass., 1950.
4. HARRIS, Z. S. Morpheme alternants in linguistic analysis. *Language*, 1942, **18**, 169–180.
5. ——. From morpheme to utterance. *Language*, 1946, **22**, 161–183.
6. ——. *Methods in structural linguistics*. Chicago: Univ. of Chicago Press, 1951.
7. HAYDEN, R. E. The relative frequency of phonemes in general American English. *Word*, 1950, **6**, 217–223.
8. HOCKETT, C. F. Problems of morphemic analysis. *Language*, 1947, **23**, 321–343.
9. ——. Age-grading and linguistic continuity. *Language*, 1950, **26**, 449–456.

[6] Psychologists have not been alone in overlooking the literature on linguistics, *cf.* (**13**).

10. JAKOBSON, R. *Kindersprache, Aphasie und allgemeine Lautgesetze.* Uppsala: Almquist and Wiksells, 1941.
11. LEOPOLD, W. F. *Speech development of a bilingual child: a linguist's record.* Evanston: Northwestern Univ. Series in the Humanities, 1939–1948, 4 Vols. (6, 11, 18, 19).
12. MILLER, G. A. Speech and language. In S. S. Stevens (Ed.), *A handbook of experimental psychology.* New York: John Wiley, 1951, pp. 789–810.
13. MOORE, O. K., & OLMSTED, D. L. Language and Professor Levi-Strauss. *American Anthropologist,* 1952, **54**, 116–119.
14. NEWMAN, E. B. The pattern of vowels and consonants in various languages. *Amer. J. Psychol.,* 1951, **64**, 369–379.
15. REED, D. W. A statistical approach to quantitative linguistic analysis. *Word,* 1949, **5**, 235–247.
16. TRAGER, G. L., & SMITH, H. L., JR. Outline of English structure. In *Studies in linguistics.* Norman, Okla. Occasional Papers No. 3.
17. VELTEN, H. V. The growth of phonemic and lexical patterns in infant language. *Language,* 1943, **19**, 281–292.
18. ZIPF, G. K. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology,* 1929, **40**, 1–95.

[MS. received August 29, 1951]

AN ANALYSIS OF MEANING [1]

BY CLYDE E. NOBLE [2]

State University of Iowa

INTRODUCTION

Intimately related to the production of research data and to the formulation of theory in psychology is the procedure of identifying and quantifying the relevant variables within its various domains. Concerning the field of verbal learning, such writers as Carr (5), Robinson (24), McGeoch (22, 23), Dashiell (8), Melton (19, 20), and Underwood (28) have emphasized the continuing need for this type of analytic research.

The analysis of the attributes of verbal material has, moreover, a more general application. The training of human perceptual and motor skills is often accompanied by verbal instructions intended to facilitate performance on such tasks. One factor which may determine the effectiveness of instructions is the nature of the verbal stimuli which are introduced. For example, particular words may vary in *meaningfulness*. Evidence reviewed in such sources as McGeoch (23), Underwood (28), and Woodworth (31) indicates this to be a potential relevant variable in verbal learning. Historically, however, there has been little agreement on the precise definition of meaning, with the result that few consistent scaling procedures have been developed with which its actual relevance may be evaluated.

The objective of the present study is a theoretical-experimental analysis of the attribute of meaning in verbal stimulus material. Defining operations will be designed in accordance with rational considerations about this concept. In addition, quantitative analysis will provide the scale values necessary to the discovery of accurate functional relationships among this and other better-known psychological variables.

AN ANALYSIS OF MEANING

The many problems of *meaning* have occupied the attention of philosophers and of psychologists for a long time. Some of these problems have been genuine, others spurious. Inasmuch as its necessary and sufficient operations are both logically possible and empirically feasible, it may be shown that at least one of these alleged issues constitutes a genuine problem. Such is the requirement of an analysis of meaning in verbal learning theory.[3]

Within the framework of Hull's (14) behavior theory this is a relatively straightforward task. Consider a stimulus element S_x, a class of conditioned responses R_1, R_2, R_3, $\cdots$ R_n, and a

[1] This report forms a part of a dissertation submitted to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree, June 1951. The author is indebted to Prof. Kenneth W. Spence for his advice and criticism. A portion of this paper was read before the Midwestern Psychological Association, April 1951.

The experimental work was conducted during the writer's temporary appointment as research psychologist with the Human Resources Research Center during the summer of 1950. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or endorsement of the Department of the Air Force.

[2] Now at the Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

[3] For a set of criteria to determine the status of a given problem, *cf.* H. M. Johnson (16). It is of historical interest that Watson's position on the meaning issue in 1925 was that the term was then unnecessary but that, should the need arise, behavioristic definition and explanation could be given (29, p. 201).

class of corresponding habit strengths H_1, H_2, H_3, $\cdot \cdot \cdot$ H_n, forming hypothetical bonds between them. Assume that these Rs have, by virtue of prior training, been connected at various times to S_x, and that, for simplicity, the Hs of these connections are severally equal. This hypothetical situation may be represented by the schema in Fig. 1, where the broken-shafted arrows denote learned connections, and where the dotted line following D denotes a continuing need or motivational state. It will be seen that this model depicts an ordinary competing response situation, in which each R has an equal probability of occurrence (R_p) following the presentation of S_x, and in which, given a wider range of observations, such phenomena as alternation, blocking, and increased reaction latency (R_t) are predictable (*cf.* **15**). Should the Hs be altered considerably in value the principle of competition would still hold, except that to refer to the habit structure under such conditions one would use the term ***hierarchy of habits***.

Since, by logical anaylsis, ***meaning*** is a relation between terms, let us define the meaningfulness of this ***situation*** as the number of Hs subsisting between S and the several Rs taken together. More specifically, the ***particular*** meanings of S_x are: H_1, H_2, H_3, $\cdot \cdot \cdot$ H_n, and different conceptual combinations of these Hs yield different ***numbers*** of meanings.

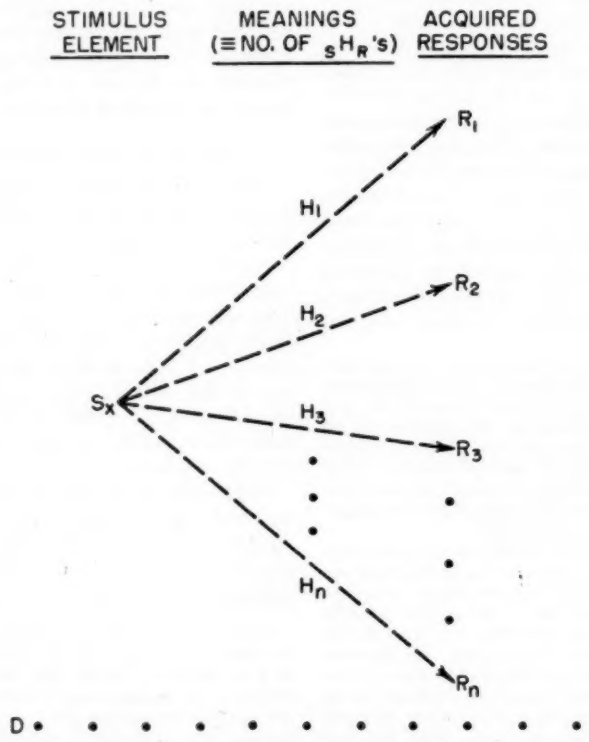

FIG. 1. Schema illustrating the development of stimulus meaning.

In this definition the author presupposes, of course, that the system *S–H–R* is isomorphic with the system *α-means-β*.

Various possible *logical* "meanings" of meaning—e.g., signification, denotation, connotation, equality, equivalence, definitional equivalence, material implication, strict implication—are not at issue here. Throughout the present analysis one must clearly and persistently distinguish between logical and psychological notions about meaning; the former class of notions is conceptual (hypothetical), the latter empirical (categorical). Further, no confusion should result from referring to relations between *S* and *R* as *psychological* (empirical) meanings, since such relations are (a) purely empirical constructs, and (b) presently to be coordinated with an operational index, *m*.

The present analysis does not assert meaning and habit strength to be identical concepts, although they have some common properties. Meanings are postulated to increase in number not as an exponential growth function of the number of *particular S-particular R* reinforcements—as *H* in Hull's theory—but rather as a simple linear function of the number of *particular S-multiple R* connections established. Now in terms of excitatory strength (*E*), where $E \equiv H \cdot D$, a specific "energized" meaning may best be regarded as an unspecified supraliminal value of effective excitatory strength ($\bar{E}$), where $\bar{E} \equiv E - I$. To strengthen $\bar{E}$ beyond the value of the limen (*L*) required for reaction evocation (*R*) may alter R_p, R_a, R_n, or R_t, but the qualitative fact that *S sometimes evokes R* is unaltered. This is the psychological connotation of the assertion: *S means R*.

From an historical point of view it may be interesting to note that the present analysis is formally analogous to certain notions advanced by the British philosophers of the 18th and 19th centuries—especially by Berkeley and by James Mill. Later, from the standpoint of introspectional psychology, Titchener taught that meaning is the conscious context which, under certain conditions, accrues to a "core" of sensory or imaginal content. This is his context "theory" of meaning, which Boring (2, pp. 185, 408) has implemented by suggesting the principle of 'accrual' to be that of association; i.e., learning. In terms of the Titchener-Boring viewpoint, then, *context* (or associated content) $\equiv$ *meaning*. Boring has also attempted to describe Titchener's doctrine in *S–R* terminology by identifying *S* with "core" and *R* with "context" (2, p. 588; 4, p. 18); hence with "meaning." However, since they can be shown to state infeasible demands (*cf.* 16), these and allied considerations —such as whether meaning is "palpable," "unconscious," or only "potential" —are of no scientific concern to modern (behavioristic) psychology.

An interesting approach to a contextual theory of meaning at a more complex level of analysis is provided by the recent paper of Miller and Selfridge (21). Defining meaningfulness in terms of dependent probabilities in successive free-association observations, these investigators have developed lists of varying orders of approximation to the structure of English.

So much for the formal definition of the concept of psychological meaning in terms of habit. Like habit, meaning is here a purely empirical construct. In order to impart significance to this notion—i.e., to relate it systematically to other constructs in behavior theory—it is first necessary to exhibit certain operations by means of which it may be given empirical verification. In other words, one must specify an empirical *index* with which the formal concept of meaning may be placed in correspondence. A few such indices are already available in the field of verbal learning.

They are defined by the various operations used for calibrating the "association values" of stimulus items. The work of Cason (6), Glaze (9), Haagen (11), Hull (13), Kreuger (18), and of Witmer (30) represent important contributions to this problem. However, most of these indices are unsatisfactory either because they involved (a) very short response intervals, (b) free association techniques, (c) relative frequency measures, or (d) because their reliabilities were not reported. Since theoretically the number of Rs is proportional to the number of supraliminal $\bar{E}$s, frequency of response is proposed as a rational index of stimulus meaning (m). Therefore, the appropriate association value for the present analysis would be denoted by the central tendency of the frequency distribution of *continued* associations given by Ss per unit time.[4]

Following these considerations an operational index of the attribute of stimulus meaning (m) was sought for each member of a list of 96 dissyllables, as indicated by the following definition:

$$m_s \equiv \frac{1}{N} \sum_{i=1}^{N} R_s, \qquad (1)$$

where $S \equiv$ verbal stimulus; $R \equiv$ unit written response; $N \equiv$ number of subjects.

Procedure

A provisional list of 120 two-syllable nouns was taken from the Thorndike-Lorge tables (26), the principal selection criterion being the G (general) frequency count. These items were drawn with the intention of representing (a) nearly all of the alphabet in the initial letters and (b) an extreme frequency-of-use range. It was hypothesized that frequency of occurrence in the written language would be highly correlated with m and therefore should provide a useful approximate ranking of the sample of stimuli to be calibrated.

The words from the Thorndike-Lorge list were supplemented by 18 artificial words which were also dissyllabic and in the form of nouns. They were selected from Dunlap's list (*cf.* 31) or invented by the author. The purpose of including these paralogs was to insure a low-m extreme on the final scale, as well as to calibrate such items empirically on a scale continuous with that of actual words.

After a number of arbitrary and systematic rejections, the final list was reduced to 96 items. This list contained approximately 20 per cent paralogs, 35 per cent infrequent items (< 1 per 4 million), and 45 per cent frequent items (> 1 per million). The last two frequency classes are defined by their Thorndike-Lorge relative frequency counts indicated in parentheses. The number 96 was selected because it was associated with a convenient maximum testing time for the prevailing military research schedules at the Human Resources Research Center. The time interval of 60 sec. was chosen because (a) preliminary tests showed that Ss reported 60 sec. to be an optimal interval, and (b) to insure a reliable time sample of the Ss' response hierarchy.

The stimulus items were administered in test booklet form with attached answer sheets. A uniform set was maintained by printing only one stimulus per page and by instructing S to return to the stimulus each time before responding anew. Furthermore, a given stim-

[4] *Continued* associations are those which are successively elicited by the same stimulus, as distinguished from *free* or *controlled* associations. The present procedure is analogous to one reported by Cattell and Bryant (7) and called, apparently by Woodworth (31), the *method of continued association*. Krueger (18) later used written responses in determining the association values of nonsense syllables, although the instructions to his Ss would seem to define their responses as *free* associations.

ulus item was reproduced on each line in an effort to reduce *S*'s inveterate tendency to free-associate. The sequence and order of presentation of the stimuli were varied by shuffling the answer sheets during the assembly of the test booklets. This device served to minimize constant errors due to fatigue, decreasing motivation, and inter-item interaction. After administering a pilot list to 15 *Ss*, in order to standardize the procedural variables, the final list of 96 items was given to a sample of 131 basic airmen from two flights undergoing routine classification testing at the Human Resources Research Center. These *Ss* were group-tested in four separate units of approximately half-flight size (about 33 men) in order to reduce inter-individual interaction. This source of variance was further reduced by varying the order and sequence of stimuli per *S*, as indicated. Two examiners tested two groups each, and the testing periods were held during the morning hours of two days one week apart. Response periods were of 60 sec. duration per stimulus, with an inter-item interval of 15 sec. Rest periods were given as follows: 5 min. at the end of the first 45 min. of testing, 10 min. at the end of the second 45 min. period, and 5 min. at the end of the next 30 min. period.

Instructions to the *Ss* were as follows:

This is a test to see how many words you can think of and write down in a short time.

You will be given a *key* word and you are to write down as many *other* words which the key word brings to mind as you can. These other words which you write down may be things, places, ideas, events, or whatever you happen to think of when you see the key word.

For example, think of the word, *KING*. Some of the words or phrases which *KING* might bring to mind are written below:

queen	Kingdom
King Cole	England
ruler	imperial
Sky-King	kingfish

Ss were then given two practice sessions, using as stimuli the words HAM and KOREA. They were permitted to use two-word phrases, slang, long words or short words, provided they were associates of the stimulus words.

Instructions regarding motivation and set were as follows:

No one is expected to fill in all the spaces on a page, but write as many words as you can which each key word calls to mind. Be sure to think back to the *key* word after each word you write down because the test is to see how many other words the key word makes you think of. A good way to do this is to repeat each key word over and over to yourself as you write.

E also gave supplementary motivating instructions during the three rest periods.

The method of recording *S*'s responses was sufficiently objective to require very little evaluation on *E*'s part. However, in terms of the analysis of meaning proposed, it was decided to set up three objective criteria for unacceptable responses. These were:

1. *Illegible responses*: $S_x \rightarrow ?$
2. *Perseverative responses*:
$$S_x \rightarrow R_1(s_1) \rightarrow R_1(s_1) \cdots$$
3. *Failures of set*: $S_x \begin{cases} \nearrow R_1 \\ \searrow R_2(s_2) \begin{cases} \nearrow R_{2_1} \\ \rightarrow R_{2_2} \\ \searrow R_{2_3} \end{cases} \end{cases}$

This last class of unacceptable responses was further classified into:

(a) *Free or tangential associations:* e.g., LEMUR→Dorothy, Hope, faith, charity. . . .

(b) *Clang or alliterative associations:* e.g., KAYSEN→caisson, Casey, casein, casement. . . .

Finally, a general rule of giving *S* the benefit of the doubt was adopted. This was occasionally necessary in the case of category 3 above, although free and

TABLE I

LIST OF DISSYLLABLE WORDS (NOUNS) IN RANK ORDER OF INCREASING MEANINGFULNESS (m) AS DEFINED BY MEAN FREQUENCY OF CONTINUED ASSOCIATIONS IN 60 SEC. ($N = 119$)

Rank	Word Number	m-Value	σ	Word	Rank	Word Number	m-Value	σ	Word
1	24	0.99	2.05	GOJEY	49	58	2.69	3.43	OVUM
2	53	1.04	1.60	NEGLAN	50	72	2.73	3.24	ROSTRUM
3	49	1.05	1.85	MEARDON	51	84	2.76	2.92	VERTEX
4	8	1.13	1.89	BYSSUS	52	5	2.80	3.27	BODICE
5	4	1.22	1.95	BALAP	53	76	2.89	3.20	TANKARD
6	86	1.22	2.17	VOLVAP	54	60	3.06	3.04	PALLOR
7	77	1.24	2.03	TAROP	55	74	3.21	2.85	SEQUENCE
8	90	1.24	2.20	XYLEM	56	1	3.34	3.34	ARGON
9	41	1.26	2.16	LATUK	57	68	3.36	3.22	RAMPART
10	66	1.26	2.01	QUIPSON	58	35	3.51	3.50	JITNEY
11	25	1.27	2.20	GOKEM	59	17	3.55	3.19	ENTRANT
12	52	1.28	1.96	NARES	60	59	3.62	3.26	PALLET
13	96	1.28	2.19	ZUMAP	61	51	3.64	3.48	NAPHTHA
14	63	1.30	1.98	POLEF	62	62	3.77	3.45	PIGMENT
15	73	1.33	2.06	SAGROLE	63	57	3.91	3.42	ORDEAL
16	55	1.34	2.37	NOSTAW	64	94	4.44	3.19	ZENITH
17	6	1.39	2.12	BODKIN	65	91	4.60	3.82	YEOMAN
18	81	1.50	2.78	ULNA	66	67	4.68	3.13	QUOTA
19	88	1.53	2.05	WELKIN	67	64	5.10	3.45	QUARRY
20	29	1.54	2.84	ICON	68	15	5.13	3.19	EFFORT
21	40	1.55	2.45	KUPOD	69	83	5.32	3.24	UNIT
22	13	1.60	2.46	DELPIN	70	18	5.33	3.46	FATIGUE
23	3	1.71	2.55	ATTAR	71	37	5.47	3.11	KEEPER
24	48	1.73	2.69	MATRIX	72	38	5.52	3.70	KENNEL
25	12	1.74	2.69	DAVIT	73	47	5.61	3.32	MALLET
26	89	1.78	2.77	WIDGEON	74	42	5.94	3.17	LEADER
27	7	1.79	2.65	BRUGEN	75	65	5.98	3.16	QUARTER
28	36	1.82	2.95	KAYSEN	76	69	5.98	3.70	REGION
29	46	1.84	2.85	MAELSTROM	77	28	6.02	3.33	HUNGER
30	79	1.84	2.95	TUMBRIL	78	95	6.15	3.05	ZERO
31	70	1.86	2.85	RENNET	79	30	6.24	3.50	INCOME
32	71	1.90	2.35	ROMPIN	80	82	6.57	3.79	UNCLE
33	22	1.95	2.55	GAMIN	81	92	6.75	4.12	YOUNGSTER
34	19	2.09	3.11	FEMUR	82	80	6.83	3.29	TYPHOON
35	45	2.09	3.42	LOZENGE	83	10	6.88	3.11	CAPTAIN
36	20	2.13	2.77	FERRULE	84	93	7.12	3.75	ZEBRA
37	75	2.14	2.75	STOMA	85	23	7.17	4.48	GARMENT
38	26	2.15	3.09	GRAPNEL	86	85	7.28	4.05	VILLAGE
39	21	2.19	3.25	FLOTSAM	87	31	7.39	3.09	INSECT
40	11	2.26	3.35	CAROM	88	34	7.58	3.69	JEWEL
41	54	2.26	2.65	NIMBUS	89	32	7.70	3.53	JELLY
42	43	2.28	3.06	LEMUR	90	27	7.91	3.86	HEAVEN
43	9	2.41	3.13	CAPSTAN	91	56	7.95	3.66	OFFICE
44	61	2.43	2.88	PERCEPT	92	87	8.12	3.67	WAGON
45	44	2.48	2.96	LICHENS	93	14	8.33	4.21	DINNER
46	33	2.54	3.53	JETSAM	94	50	8.98	4.27	MONEY
47	16	2.59	3.08	ENDIVE	95	2	9.43	4.30	ARMY
48	78	2.63	3.04	TARTAN	96	39	9.61	4.30	KITCHEN

clang associations were usually easily identified by the three scorers. Of the original 131 *Ss* tested, 12 protocols were rejected for persistent violations of the criteria cited. This brought the effective sample to 119 *Ss*.

Results

The index of meaning (m) of a particular stimulus was defined in equation [1] as the grand mean number of (acceptable) written responses given by all *Ss* within a 60 sec. period. Therefore, the scale values of the stimuli were determined directly by the average response frequencies of the 119 *Ss*. These m-values with the σ's of their distributions are shown in Table I, ranging in rank order from dissyllables of low response-evocation value (e.g., No. 1: GOJEY) to those of high response-evocation value (e.g., No. 96: KITCHEN). The empirical range is from 0.99 to 9.61. It will be noted that there is no discrete gap between the paralog items and the actual words. In fact, there are a few actual words low on the scale (e.g., No. 4: BYSSUS), while one paralog (No. 32: ROMPIN) appears at the third-way point. It is also to be noted that response variability tends to increase with increasing m-value.

It was found that the m-values of particular items were quite stable from group to group. Intercorrelations of the four sets of mean m-values per word were carried out among all six combinations of groups. These Pearsonian r-values appear in Table II. All are significantly different from zero ($P < 0.01$). Since the sampling distribution of r is skewed for large values, Fisher's Z-transformation was used to estimate the mean intergroup reliability coefficient of the m-scale: $r_{mm} = 0.975$. It may be pointed out that a between-groups reliability coefficient [5] is the appropriate statistic to compute in this case since it was *E*'s aim to determine the consistency of *different response samples* to the *same stimuli*. A more conventional reliability coefficient —such as one defined by the test-retest, split-half, or the alternate form procedure—would not have evaluated this particular relationship.

TABLE II

INTERGROUP RELIABILITY COEFFICIENTS (r) FOR m-SCALE BASED ON MEAN m-VALUES FOR FOUR GROUPS OF *Ss*

Group	I	II	III	IV	n
I	—	—	—	—	27
II	.98	—	—	—	30
III	.98	.98	—	—	30
IV	.96	.97	.98	—	32

N = 119.

Of some interest in this investigation was the extent to which the assumptions of the product-moment correlational method were met. The family of response frequency distributions associated with the stimulus items exhibited skewness at the low-m extreme of the scale, but throughout the central and upper portions they were approximately symmetrical. Hence, means were retained as measures of central tendency. When the six intercorrelations of the m-scale were plotted, it was found that the requirements of linearity of regression and of homoscedasticity were reasonably satisfied.

DISCUSSION

Meaning. After reviewing the analysis of meaning, one properly may ask whether responses may acquire meanings also, or whether only stimuli do so. As has been indicated, meanings are considered relations between *Ss* and *Rs*.[6]

[5] Rather than a "reliability" coefficient, some might prefer to call this statistic a coefficient of "objectivity," "agreement," or of "consistency."

[6] Not all writers would agree to construe meaning thus (*cf.* 2, 3, 4, 17, 27). Some of the divergences of opinion are reflected in the following successive quotations from Boring's

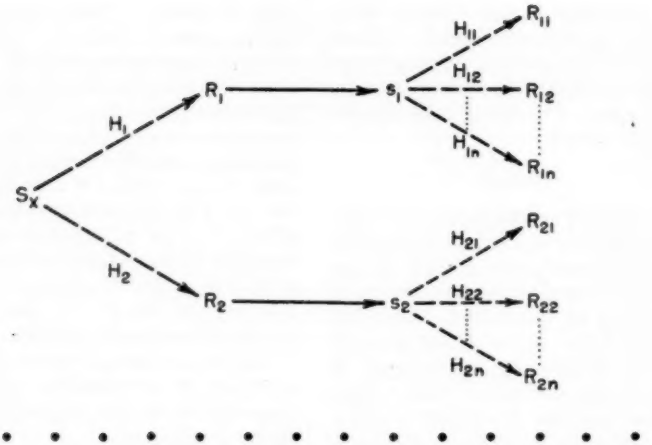

FIG. 2. Schema illustrating the development of more complex degrees of meaning.

It is for convenience that, instead of referring to *m*, one speaks of "the meaning of a stimulus" or of "stimulus meaning," just as one speaks of "reinforcing a conditioned response," when more precisely it is the *S–R* connection, or *H*, which strictly is reinforced. To speak of *stimulus* meaning is to imply an asymmetry in the empirical meaning relation. Indeed, common linguistic usage seems to concur that meaning, like causation, be regarded as a property of stimuli rather than of responses.[7]

Under certain conditions, however, it is appropriate to refer loosely to the meaning of a *response*. One such condition would arise if analysis should indicate that a certain class of responses results in proprioceptive stimulation (*s*) to an organism. This is the well-known response-produced stimulus situation, or feedback mechanism. Kinesthesis is perhaps the best example. The descriptive schema outlined in the section on meaning applies equally well, as shown in Fig. 2. Here the situation becomes more complex, but no different qualitatively. *H*s develop between initial *R* ($\rightarrow s$)s and subsequent *R*s, thereby permitting more complex degrees of development of meaning.

The serial anticipation learning situation, for example, fits this schema well in that each successive response functions as a stimulus to the next response in the series, while the verbal stimuli appearing serially in the memory drum aperture serve as secondary reinforcing agents. This type of analysis also suggests an explanatory approach for the relationships Miller and Selfridge (21) have found between reproductive recall and contextual dependencies. The work of Thorndike (25) on "belonging" may be similarly regarded.

The result of this analysis of meaning and its incorporation into learning theory is twofold: (a) it establishes

1933 text (3): "The correlation between stimulus and ultimate response *is* a meaning . . ." (p. 153); ". . . the data of consciousness are meanings" (p. 222); ". . . a meaning is a relation" (p. 222); "Meaning is response" (p. 223).

[7] This is not to be interpreted as bearing any relation to the Gestalt doctrine of meaning as immediately given in the external world (*cf.* 17), but simply as a clarification of a familiar idiom.

a highly reliable, unequivocal new attribute of variation in learning, and (b) it provides an operational basis for explicating the common-sense notion of "meaning." Thus, if one were to ask a layman what he intended by saying that "home" to him *means:* "family, spouse, children, friends, love," etc., he would doubtless reply, "I think of these things when 'home' is mentioned." It is a simple matter to fit this statement to the model proposed in Fig. 1. A learning theorist would explain that to the auditory (or visual) *S home,* these various verbal *R*s have become conditioned during our imaginary layman's previous experience, and that under appropriate conditions (e.g., adequate *D* level, sufficient reinforcements to each *L*) these *R*s are elicited. The meaning of *S* subsists in the *H*s developed to it —nothing more.[8] A neutral *S*, by the present definitions, is meaning*less*; an *S* conditioned to twenty *R*s is more meaning*ful* (i.e., has more meanings) than is one conditioned to ten, and so on. Speaking quite non-technically, meanings are habits. And as more habits accrue to a particular stimulus situation, so does its meaningfulness increase.

In the final analysis the index *m*, like *H*, emerges as a statistical concept. It is a function of the number of particular *S*-multiple *R* connections which are formed, and it enjoys an existence independent of the cited operations no more than does any other empirical construct.

In view of its theoretical status and exceptional reliability, the *m*-scale is regarded with especial interest for research in verbal and in perceptual-motor learning. Using the *m*-scale, research may now be directed toward the solution of such current problems as the relationship between difficulty and meaningfulness, the effect of meaningfulness upon reminiscence, and the rôle of verbal instructions in the acquisition of motor skills.

SUMMARY AND CONCLUSIONS

This paper has presented a theoretical-experimental analysis of the attribute of meaning in verbal stimulus material. A word list of 96 dissyllables consisting of nouns and paralogs was presented to a sample of 119 USAF recruits in order to establish a quantitative scale for this attribute. The results of this analysis were as follows:

1. Meaning was formally defined as a relation between *S* and *R*. It was coordinated with Hull's theoretical construct *H* by postulating that meanings increase as a simple linear function of the number of *S*-multiple *R* connections acquired in a particular organism's history.

2. An index of stimulus meaning (*m*) was operationally defined in terms of the mean frequency of continued written associations made by subjects within a 60-sec. time interval.

3. A psychological performance scale of *m*-values was developed which exhibited a range extending from 0.99 to 9.61. The mean product-moment intergroup reliability coefficient was $r_{mm} = 0.975$.

4. The significance of these findings for human learning theory was discussed.

REFERENCES

1. BERGMANN, G., & SPENCE, K. W. The logic of psychophysical measurement. PSYCHOL. REV., 1944, 51, 1–24.

[8] In view of the fact that the class of responses defining the *m*-scale represents a set of individual-difference variables, one should properly denote the meaning relation in this case by the symbol *U* rather than by *H*. From a genetic standpoint, certain learning theorists preserve a distinction between innate *or* previously-acquired connections (*sUr*'s) and laboratory-established connections (*sHr*'s). However, in Hull's theory both *U* and *H* are considered habit or associational concepts, hence the proposed coordination with meaning is unimpaired. A similar argument holds for conditioned generalized connections (*sH̄r*'s).

2. BORING, E. G. *A history of experimental psychology*. New York: D. Appleton-Century Co., 1929.
3. ——. *The physical dimensions of consciousness*. New York: D. Appleton-Century, 1933.
4. ——. *Sensation and perception in the history of experimental psychology*. New York: D. Appleton-Century, 1942.
5. CARR, H. A. *Psychology: A study of mental activity*. New York: Longmans, Green, 1925.
6. CASON, H. Specific serial learning: a study of backward association. *J. exp. Psychol.*, 1926, **9**, 195–227.
7. CATTELL, J. McK., & BRYANT, S. Mental association investigated by experiment. *Mind*, 1889, **14**, 230–250.
8. DASHIELL, J. F. A neglected fourth dimension to psychological research. PSYCHOL. REV., 1940, **47**, 289–305.
9. GLAZE, J. A. The association value of non-sense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
10. GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
11. HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.
12. HEBB, D. O. *The organization of behavior*. New York: John Wiley & Sons, Inc., 1949.
13. HULL, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.
14. ——. *Principles of behavior*. New York: Appleton-Century, 1943.
15. ——. *Elementary theory of individual behavior*. (In preparation.)
16. JOHNSON, H. M. Are psychophysical problems genuine or spurious? *Amer. J. Psychol.*, 1945, **58**, 189–211.
17. KÖHLER, W. *The place of value in a world of facts*. New York: Liveright, 1938.
18. KRUEGER, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, **17**, 145–153.
19. MELTON, A. W. The methodology of experimental studies of human learning and retention: I. The functions of a methodology and the available criteria for evaluating different experimental methods. *Psychol. Bull.*, 1936, **33**, 305–394.
20. ——. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1950, 668–690.
21. MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.
22. McGEOCH, J. A. The vertical dimensions of mind. PSYCHOL. REV., 1936, **43**, 107–129.
23. ——. *The psychology of human learning*. New York: Longmans, Green, & Co., 1942.
24. ROBINSON, E. S. *Association theory today: An essay in systematic psychology*. New York: Century, 1932.
25. THORNDIKE, E. L. *Human learning*. New York: Century, 1931.
26. ——, & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia Univ. Press, 1944.
27. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Century Co., 1932.
28. UNDERWOOD, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, Inc., 1949.
29. WATSON, J. B. *Behaviorism*. New York: Norton & Co., 1925, 2nd Ed.
30. WITMER, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–359.
31. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

[MS. received September 4, 1951]

THE ROLE OF OBSERVING RESPONSES IN DISCRIMINATION LEARNING [1]

PART I

BY L. BENJAMIN WYCKOFF, JR.
University of Wisconsin

Theorists in the area of discrimination learning have often had occasion to refer to a set or predisposition of S to learn differential responses to a particular pair of stimuli. Such a predisposition has often been attributed to some reaction of S such as an attending response, orienting response, perceiving response, sensory organizational activity, etc. To implement the discussion of the role of such reactions in discrimination learning we shall adopt the term "observing response" (R_o) to refer to any response which results in exposure to the pair of discriminative stimuli involved. The probability of occurrence of an observing response will be denoted by p_o. These responses are to be distinguished from the responses upon which reinforcement is based; that is, running, turning right or left, lever pressing, etc., which, for convenience, we shall term "effective responses."

Spence (19) has proposed a theory of discrimination which is specifically intended to deal with situations where no observing response is required of S, that is to say, to situations in which S is certain to be exposed to the discriminative stimuli on each trial or prior to each effective response ($p_o = 1$). The fact that in some discrimination experiments this condition has not been satisfied has become an issue in the literature, largely because it became necessary to delimit clearly the situations to which Spence's theory is intended to apply.

Spence's theory of discrimination states that stimulus-response connections are strengthened or weakened during discrimination training in essentially the same way as these changes would occur during conditioning or extinction. When a response is reinforced the connections between it and all aspects of the stimulus situation impinging on S at the time the response occurred will be strengthened. These connections will be weakened when the response is not reinforced. Certain implications of this theory were questioned by Krechevsky (11) and other theorists, and became the subject matter of the "continuity-discontinuity" controversy. This material has been reviewed a number of times (2, 5) and need not be repeated in detail here. One aspect of the controversy is pertinent to the present discussion. Krechevsky (12) presented experimental findings which indicated that rats learned nothing with respect to two stimulus patterns during the first 20 trials of a discrimination experiment even though they were systematically reinforced for approaching a particular pattern during this interval. Failure to learn was established by showing a lack of interference when *Ss* were tested on a reversed discrimination. These findings were in apparent disagreement with the data obtained by McCulloch and Pratt (13) in a similar experiment in which differing weights were used as discriminative stimuli. Here interference was obtained, indicating that some cumulative learning had occurred in the early portion of the experiment.

In interpreting these results, Spence (20, p. 277) argued that the stimuli (patterns) used by Krechevsky were not sufficiently

[1] This paper is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Department of Psychology, Indiana University. The writer wishes to express his appreciation to Dr. C. J. Burke for his invaluable guidance and stimulation.

conspicuous to provide a legitimate test of his theory. He suggested that *S*s had not learned to orient toward the stimuli within the first 20 trails. He points out that in such cases, ". . . the animal must learn to orient and fixate its head and eyes so as to receive the critical stimuli." He then suggests a way in which this learning may occur. "These reactions are learned . . . because they are followed within a short temporal interval by the final goal response."

This interpretation was put to an experimental test by Ehrenfreund (5). In his experiment the likelihood of *S*'s receiving the critical stimuli was manipulated by changing the position of the stimuli (upright and inverted triangles) with respect to the landing platform of a jumping stand. The design of the experiment was essentially the same as Krechevsky's. The results conform to Spence's interpretation. When the stimuli were placed relatively high, no learning occurred within the first 40 trials, whereas when the stimuli were placed closer to the landing platform learning did occur. Learning was again measured in terms of interference in the learning of a subsequent reversed discrimination.

The analysis of discrimination situations in which some observing response is required is of interest for several reasons. First, discrimination learning in situations other than laboratory experiments, such as human learning in the course of every day events, is largely of this kind. Secondly, even in the most closely controlled laboratory experiments it is seldom, if ever, possible to say with certainty that *S* is exposed to the discriminative stimuli prior to each effective response. In the case of pattern discriminations it has been demonstrated by Ehrenfreund (5) that relatively small differences in the position of the discriminative stimuli will effect discrimination learning, indicating that relatively precise fixation of the stimulus is required.

In the present paper an attempt will be made to develop a more extensive theory of discrimination which will include situations in which some observing response (hereafter referred to as R_o) is required before *S* is exposed to the discriminative stimuli. An example of such a situation would be an experiment in which stimulus cards were placed overhead. In this case the response of raising the head would be the R_o).

If we accept the notion that changes in p_o can be accounted for within the framework of reinforcement learning theory, it should be possible to devise a theory of discrimination which will include those cases where some R_o is necessary. The purpose of this paper is to outline such a theory. We shall see that by analyzing discrimination learning in this way it will be possible to account for stimulus generalization and also changes in generalization during discrimination learning without postulating any direct interaction between stimuli. Several hypotheses will be derived from this theory which have been tested in an experiment by the author presented in detail elsewhere (22). Finally we shall outline a way in which the present theory can be integrated with existing quantitative theories of conditioning and extinction to form a quantitative theory of discrimination.

To simplify this discussion let us consider a hypothetical experiment using a situation similar to that used by Wilcoxon, Hays, and Hull (21), and later used by Hull (10) for a discrimination experiment. In this experiment a rat was placed in a small compartment with a single exit through a door into a goal compartment. A measure of the latency of the response of running through this door was obtained. The discriminative stimuli consisted of a black or a white door, either one of which was

present on each trial. During discrimination training the running response was reinforced with food when one color was present, whereas reinforcement was withheld when the other color was present. Each stimulus was present on an average of 50 per cent of the trials.

For purposes of the present discussion let us consider a slightly different situation in which the discriminative stimuli are placed overhead rather than directly in front of *S*. In this case an observing response, raising the head, will be necessary if *S* is to be exposed to the discriminative stimuli. On each trial, when *S* is placed in the apparatus, there will be a certain probability that the R_o of looking up will occur. When R_o does occur *S* will be exposed either to a black or a white card. When the R_o fails to occur *S* will not be exposed to either card, but rather to a neutral population of stimuli (walls, floor, etc.). Note that in this situation *S* does not improve its chances of ultimate reinforcement by making the R_o. The food is placed in the goal compartment whenever the white card is present whether *S* actually looks up or not. In a sense then, *S* gains only information by making the R_o.

We are now in a position to examine the relation between observing responses and stimulus generalization. In general it is apparent that if p_o has a low value, *S* will seldom be exposed to the discriminative stimuli (the black and white cards). *S* therefore, will have minimum opportunity to learn discrimination or to manifest any discrimination already learned. On the other hand, if p_o has a high value, the opportunity to learn or manifest discrimination will be large.

Stimulus generalization between two stimuli is usually defined either in terms of *S*'s tendency to respond similarly to the two stimuli, or in terms of failure to learn differential responses readily. Thus we can see that stimulus generalization will decrease as p_o increases.

If we assume that p_o changes as a result of learning processes we can see that these changes would give rise to changes in generalization between the stimuli involved. More specifically, if we assume that p_o will increase during discrimination learning (differential reinforcement), generalization between the discriminative stimuli will decrease. Similarly, we might assume that p_o will decrease if we introduce a procedure in which the subject is reinforced equally often in the presence of either stimulus (non-differential reinforcement). This decrease in p_o would give rise to an increase in generalization between the stimuli.

In the case of the hypothetical experiment suggested above, generalization will be shown in a "crossover" effect between positive and negative trials. Reinforcements on positive trials (positive stimulus card present but not necessarily observed) will tend to strengthen the effective response on negative trials, while unreinforced responses on negative trials will tend to weaken the effective response on positive trials. If *S*'s tendency to look up increases during differential reinforcement, this "crossover" effect will decrease. If during non-differential reinforcement the tendency to look up decreases, the "crossover" effect will increase.

It should be emphasized that these statements regarding increases and decreases in p_o are, at this point, assumptions which may or may not be true in a particular experimental situation. We shall present experimental findings which suggest that these assumptions are quite generally true below.

In the above discussion we have considered the effects of R_o on discrimination and generalization. At this point we turn our attention to the problem of accounting for changes in p_o within the framework of reinforcement learning theory. Our problem will be to identify possible reinforcing conditions which may account for increases in p_o during differential reinforcement.

First we note that, by definition, the observing response results in exposure to a pair of discriminative stimuli. If exposure to these stimuli is in some way reinforcing, we shall expect p_o to increase or remain high. The problem at hand is to show how exposure to discriminative stimuli may have a reinforcing effect under the condition of differential reinforcement, while the same stimuli do not have this effect under the condition of non-differential reinforcement. Reinforcement theory provides two ways of accounting for this reinforcing effect.

The first method is the mechanism suggested by Spence when he states that observing responses are learned "because they are followed within a short temporal interval by the final goal response" (19). This mechanism will operate in experiments such as a "jumping stand" experiment, in which exposure to discriminative stimuli may serve to increase the probability of prompt reinforcement, that is to say, the probability of the "correct" jump may be increased. Spence offered this suggestion in relation to a jumping stand experiment.

The second method of accounting for the reinforcing effect is by appeal to the principles of secondary reinforcement. Here we suggest that the discriminative stimuli themselves take on secondary reinforcing value during the course of discrimination learning. It has been demonstrated that an originally neutral stimulus which accompanies reinforcement may acquire secondary reinforcing properties. That is, it may serve to strengthen a response upon which it is made contingent. Skinner (18, p. 246) has demonstrated that whenever a stimulus becomes a discriminative stimulus for some response in a chain leading ultimately to reinforcement, this stimulus will serve as a secondary reinforcing stimulus. The conditions necessary for the formation of secondary reinforcing properties are further considered by Notterman (16), Schoenfeld *et al.* (17) and Dinsmoor (4). They point out that in all cases where secondary reinforcement has been demonstrated, the conditions were also appropriate for the establishment of the stimulus in question as a discriminative stimulus. They suggest that this may be a necessary (as well as sufficient) condition for the establishment of secondary reinforcing properties. In the present formulation it is apparent that the positive stimulus is presented in the appropriate temporal position to become both a discriminative stimulus (for the effective response) and a secondary reinforcing stimulus (for the observing response).

This mechanism may operate in any situation whatever where an R_o is involved, since it is a defining characteristic of the R_o that it leads to exposure to discriminative stimuli. Specifically it should apply to the hypothetical experiment suggested above. Here the effective response (running) will always be reinforced when S is exposed to the white card. Hence the white card could be expected to acquire secondary reinforcing value. It is not sufficient to show simply that the positive stimulus will acquire secondary reinforcing value. We must also consider two other factors. First, R_o results in exposure to the positive stimulus only 50 per cent of the time. It results in exposure to the negative stimulus the other 50 per cent. Second, the running response is reinforced sometimes

when S is exposed to the neutral stimulus population, since, on positive trials, the running response is reinforced even though S does not look up. The effective response is reinforced most consistently when S is exposed to the positive stimulus. Therefore, it is still plausible to postulate that the intermittent exposure to the positive and negative stimuli will have a net reinforcing effect on R_o.

It is true of both of these mechanisms that, before any increase in p_o can be expected to occur, S must learn differential effective responses, that is to say, S must learn to respond differently to the two discriminative stimuli. In the case of the "jumping stand" experiment, if S does not have differential jumping tendencies toward the discriminative stimuli, the probability of reinforcement will always be 50 per cent, and will not be improved by the occurrence of R_o.

When we apply the secondary reinforcement principle we can see that the positive stimulus must appear in the proper temporal relation to reinforcement a number of times before this stimulus will acquire secondary reinforcing properties. In terms of Notterman, Schoenfeld, and Dinsmoor's interpretation it will be necessary for S to learn differential effective responses to the discriminative stimuli before secondary reinforcing properties are acquired by these stimuli.

In view of these considerations we introduce the following general hypothesis: Exposure to discriminative stimuli will have a reinforcing effect on the observing response to the extent that S has learned to respond differently to the two discriminative stimuli.

Hereafter we shall refer to the magnitude of the difference between *S*s' tendencies to respond to the two discriminative stimuli as the "degree of discrimination."

Earlier it was pointed out that the probability of occurrence of R_o is one of the factors determining the rate of formation of discrimination. According to the present hypothesis the opposite relationship is also true. The resulting picture is one of a circular interrelationship, in which R_o affects the formation of discrimination because of its effect on exposure to discriminative stimuli, while the degree of discrimination affects R_o through another mechanism involving either secondary reinforcement or changes in the probability of reinforcement.

We now present four propositions which are implied by this general hypothesis. The hypothesis was formulated partly on the basis of experimental evidence already available, which suggested that these propositions were true (**22**). At present we shall consider them as specific hypotheses. The first two of these have already been introduced as assumptions.

1. p_o will increase (or remain high) under conditions of differential reinforcement.
2. p_o will decrease (or remain low) under conditions of non-differential reinforcement.

It is apparent that these hypotheses are consistent with the general hypothesis since the degree of discrimination will tend to increase (or remain high) under differential reinforcement, while it will tend to decrease (or remain low) under nondifferential reinforcement. In other words, S will learn to respond differently to the two stimuli under differential reinforcement, but will learn to respond in the same way to them under nondifferential reinforcement. Additional

hypotheses of interest can be derived from this general hypothesis.

3. When a well established discrimination is reversed p_o will decrease temporarily and then return to a high value.

We shall expect this change in p_o because, following a reversal, the degree of discrimination will decrease as the original discrimination vanishes. It will then increase as the new discrimination is formed.

4. If at some point in an experiment the degree of discrimination is low and at the same time p_o is low (but greater than zero), we shall expect the formation of discrimination to be retarded for some interval, but finally to occur quite rapidly.

This hypothesis arises from the fact that increases in the degree of discrimination, and increases in p_o, are dependent upon each other. Early in the process S will be exposed to the discriminative stimuli only a small proportion of the time and hence the degree of discrimination cannot increase rapidly. At the same time p_o will not increase because of the low degree of discrimination. Then, as the degree of discrimination becomes sufficiently great to bring about an increase in p_o the entire learning process will be accelerated.

Krechevsky (11) presents data obtained in discrimination experiments in a jumping stand situation which correspond in some respects to the predictions of the present formulation. Curves for individual *S*s show relatively abrupt discrimination formation. In general the curves also show a slight improvement in discrimination prior to the abrupt change. A curve presented for discrimination reversal shows a rapid decrease in the degree of discrimination to a chance level, followed by an interval during which improvement was much less rapid. Finally the process accelerated as the reversed discrimination formed. Krechevsky also noted that during the interval while S was responding approximately according to chance with respect to the discriminative stimuli, he showed a strong position preference. These findings are in complete agreement with hypotheses 3 and 4 in the present formulation.

The four hypotheses presented so far were tested in an experiment by the writer (22) which is presented in detail elsewhere. In this experiment direct measures of an R_o were obtained during differential reinforcement, nondifferential reinforcement and during discrimination reversal. Pigeons were used in a Skinner-box situation in which the effective response was striking a single translucent key. The discriminative stimuli were colored lights (red and green) projected on the back of the key one at a time. The colored lights were withheld and the key was lighted white until the R_o occurred. The R_o consisted of stepping on a pedal on the floor of the compartment. The reasons for using this response as an observing response are discussed in detail elsewhere (22). Here it will suffice to say that this response falls within our definition of an observing response in that it resulted in exposure to the discriminative stimuli. As in the case of the hypothetical experiment discussed above, the observing response had no effect on the probability of reinforcement at any given moment.

All of the above hypotheses were supported by the results of this experiment. Concerning the first three hypotheses, p_o was higher under differential reinforcement than under nondifferential reinforcement. When *S*s were shifted from differential to nondifferential reinforcement a marked decrease in p_o occurred. All of these differences were significant at a 5 per cent level of confidence or better.

The fourth hypothesis does not ap-

ply unless at some point in the experiment the degree of discrimination and p_o are both low. This condition was not satisfied consistently since the operant (or base) level of the pedal response turned out to be relatively high for *S*s. However, in several cases this condition was satisfied and in these cases the results conformed to the hypothesis.

We can now illustrate some ways in which this theory might be useful in interpreting behavior in other experiments.

1. If this theory is applied to situations in which more than one pair of discriminative stimuli is involved we can make some predictions regarding changes in the readiness of *S* to form discriminations based on some particular pair of stimuli.

2. It has been demonstrated that when a discrimination is reversed repeatedly *S*s tend to learn the reversed discrimination more and more rapidly (**15, 8**). According to the present theory, during discrimination reversal the observing response is partially extinguished and reconditioned. Thus, during repeated reversals, the R_o is, in effect, reinforced intermittently. Studies of intermittent reinforcement have indicated that when a response is intermittently extinguished and reconditioned, the strength of the response tends to attain a relatively constant high value (**18**). On the first reversal p_o might drop to a low value, and recover slowly, but with repeated reversals we would expect this drop to become less prominent, and finally, p_o would remain high throughout the reversal. It is apparent that if p_o remained high, a reversed discrimination would be learned more rapidly than otherwise.

In the preceding discussion we have examined some of the ways in which discrimination learning may be affected when some observing response is required of *S*. We shall now derive some quantitative statements to supplement the above analysis. We shall attempt to set down the relationships involved in such a way that the present theory can be readily integrated into existing quantitative theories of learning such as Hull's (**9**), Estes' (**6**) or Bush and Mosteller's (**3**). The potential applications of this development could proceed along two different lines.

First, we could attempt to state the relationships between observing responses and measurable aspects of the effective responses in such a way that p_o could be estimated in situations where direct measurement of R_o is not feasible. This might be the case, for example, if the R_o involved focusing of the eye. If we apply the present development in this way, p_o would become an intervening variable, which could be used to account for and predict behavior in situations where (1) the apparent generalization between stimuli changes, or (2) where the ease of formation of discrimination changes as a function of training. Berlyne (**1**) suggests that "attention" be treated in a similar way.

Secondly, we could predict discrimination learning functions by adopting some set of assumptions regarding the component learning processes involved. These assumptions could be adopted from some existing theory which treats the simpler processes of conditioning and extinction. The main obstacle to this endeavor at the moment is the absence of any quantitative function for predicting changes in p_o. However, we shall be able to set down the relationships involved in such a way that any acceptable function can immediately be inserted.

QUANTITATIVE ANALYSIS

For purposes of this analysis let us return to consideration of the hypothetical experiment discussed above. There it was pointed out that we must take into consideration three different stimulus populations which may effect *S*s' behavior. We shall adopt the following notation to represent these stimuli. Let S_1 represent the stimulus population to which *S* is exposed on trials when the R_o occurs and when the positive stimulus card (white) is present, S_2 represent the stimulus population on trials when the R_o occurs and when the negative stimulus card (black) is present, and S_3 the stimulus population to which *S* is exposed when the R_o fails to occur.

In this analysis we shall use the symbol p to represent the probability of occurrence of the effective response at any given moment during a trial. This variable can be related to the variable of response latency as follows. Estes (6) has shown that if a response can be expected to occur with a given probability at any momnet during a trial, the mean latency of the response will be proportional to the reciprocal of the probability; that is to say, $L = k/p$, where L is the mean latency, p the probability, and k a constant of proportionality which will depend on the units of measurement used. In the present case we must consider the probability of occurrence of the effective response for each of three stimulus populations. Let us adopt the symbols p_1, p_2, and p_3 to represent the probability of occurrence of the effective response when *S* is exposed to S_1, S_2, and S_3, respectively. We shall also wish to refer to the net probability of occurrence of the effective response on a given trial, taking into account that *S* may be exposed to different stimuli during the trial depending on the occurrence or non-occurrence of the R_o. We shall use the symbols p_+ and p_- to represent the net probability on trials when the positive or negative stimuli are present.

To summarize:

S_1 = the population of stimuli to which *S* is exposed if (1) the positive stimulus is present and (2) the R_o occurs.

S_2 = the population of stimuli to which *S* is exposed if (1) the negative stimulus is present and (2) the R_o occurs.

S_3 = the population of stimuli to which *S* is exposed if the observing response fails to occur.

p = the probability that the effective response will occur at any given moment during a trial ($= k/L$)

p_1 = the value of p when *S* is exposed to S_1

p_2 = the value of p when *S* is exposed to S_2

p_3 = the value of p when *S* is exposed to S_3

p_+ = the net value of p for a trial on which the positive stimulus is present

p_- = the net value of p for a trial on which the negative stimulus is present

p_o = the probability of occurrence of R_o at any given moment during a trial.

We shall now express certain functional relationships among these variables. First we shall express p_+ and p_- as two functions of the variables p_1, p_2, p_3, and p_o. p_+ and p_- are variables which can be evaluated from experimental measures, such as latency of the effective response, without reference to direct measures of R_o. They correspond to the measures of response tendency usually obtained in discrimination experiments.

However, in the present framework p_+ and p_- are assumed to be the net result of the operation of the variables p_1, p_2, p_3, and p_o. Our task will be to express this dependence as a pair of functional relationships. This can be done as follows.

Consider a selected moment during a positive trial. At this moment S will be exposed to either S_1, with a probability of p_o, or to S_3, with a probability of $(1 - p_o)$. If S is exposed to S_1 he will make the effective response with a proabability of p_1. If the effective response and R_o are independent of each other the probability that both R_o and the effective response will occur will be the product p_1p_o. If S is exposed to S_3 he will make the effective response with a probability of p_3, and the probability that both will occur will be the product $(1 - p_o)p_3$. The total probability that the effective response will occur at this moment will be the sum of these products. Thus:

$$p_+ = p_op_1 + (1 - p_o)\, p_3. \quad (1)$$

By exactly parallel reasoning with respect to a selected moment during a negative trial we obtain:

$$p_- = p_op_2 + (1 - p_o)\, p_3. \quad (2)$$

The next step will be to derive expressions for predicting the values of p_1, p_2, and p_3. The reinforcement contingencies for the effective response in the presence of S_1, S_2, and S_3 can be readily ascertained. It will be possible to predict changes in the values of p_1, p_2, and p_3 on the basis of learning functions for the simpler processes of conditioning and extinction if we assume that learning with respect to each of these stimuli, proceeds independently of learning with respect to the others. This assumption implies that interaction between stimuli will have a negligible effect. However, in making this assumption we do not forfeit the ability to handle stimulus generalization within the present framework, since, as we have already pointed out, stimulus generalization can be accounted for without postulating any such direct interaction.

In the present paper we do not adopt a particular set of functions for conditioning and extinction, but attempt to set down the relationships in such a way that any acceptable set of functions can be immediately inserted.

The assumption of "negligible direct interaction" implies that changes in the probability of occurrence of the effective response with respect to a particular stimulus population S_i ($i = 1$, 2, or 3) will occur only during the time in which S is exposed to S_i, and that the rate of change with respect to time will depend on:

1. Whether or not the effective response is reinforced.
2. The value of p_i at the time.

If we let r_i represent the proportion of the time during which S is exposed to S_i, the rate of change of p_i can be approximated by two functions as follows:

$$dp_i/dt = r_if_e(p_i) \quad (3)$$

if the effective response is reinforced, and

$$dp_i/dt = r_if_o(p_i) \quad (4)$$

if the effective response is not reinforced.

The functions f_e and f_o represent any acceptable set of analytic functions which approximate the rate of change of probability of occurrence of an effective response during conditioning and extinction, respectively. It will be noted that if we assign a value of 1 to r we will obtain expressions for simple cases of conditioning

or extinction. In the present model the values of r_i can be expressed as functions of p_o as follows. The positive and negative stimuli are each to be present 50 per cent of the time. During this time the subject will be exposed to S_1 or S_2 with a probability of p_o. Hence:

$$r_1 = r_2 = .5p_o.$$

S will be exposed to S_3 with a probability of $(1 - p_o)$. Hence:

$$r_3 = (1 - p_o).$$

We also know that all effective responses in the presence of S_1 are reinforced, effective responses in the presence of S_2 are not reinforced, and effective responses in the presence of S_3 are reinforced an average of one-half of the time. Using the above values of r and appropriate functions for reinforced and non-reinforced responses we obtain:

$$dp_1/dt = .5p_o f_c(p_1) \quad (5)$$

$$dp_2/dt = .5p_o f_e(p_2) \quad (6)$$

$$dp_3/dt = .5(1 - p_o)f_c(p_3) + .5(1 - p_o)f_e(p_3). \quad (7)$$

We can now outline the steps which would be necessary to predict p_+ and p_- (measurable aspects of effective responses) if we can predict the values of p_o as a function of time. Such a function could be derived empirically or through some theoretical statement regarding the factors which bring about changes in p_o. If p_o can be expressed as a function of time we can rewrite equations 5, 6, and 7 to obtain expressions involving only dp_i, dt, p_i and t. If these differential equations can be solved we will obtain $p_i = f_i(t)$. Thus we can obtain values of p_1, p_2, p_3, and p_o for any point in time. These values can be substituted in equations 1 and 2 to give the desired prediction of p_+ and p_-.

On the other hand if we wish to estimate the values of p_o from known values of p_+ and p_-, we can proceed as follows.

Equations 1 and 2 state:

$$p_+ = p_o p_1 + (1 - p_o)p_3, \quad (1)$$

$$p_- = p_o p_2 + (1 - p_o)p_3. \quad (2)$$

Differentiating with respect to time we obtain:

$$dp_+/dt = p_o(dp_1/dt) + p_1(dp_o/dt) + (1-p_o)(dp_3/dt) - p_3(dp_o/dt), \quad (8)$$

$$dp_-/dt = p_o(dp_2/dt) + p_2(dp_o/dt) + (1-p_o)(dp_3/dt) - p_3(dp_o/dt). \quad (9)$$

Substituting values for dp_1/dt, dp_2/dt and dp_3/dt from equations 5, 6, and 7 and rearranging terms we obtain:

$$dp_+/dt = .5p_o^2 f_c(p_1) + .5(1-p_o)^2[f_c(p_3) + f_e(p_3)] + (p_1 - p_3)(dp_o/dt), \quad (10)$$

$$dp_-/dt = .5p_o^2 f_e(p_2) + .5(1-p_o)^2[f_c(p_3) + f_e(p_3)] + (p_2 - p_3)(dp_o/dt). \quad (11)$$

Equations 1, 2, 10, and 11 represent four simultaneous equations. By combining these equations we can express p_1, p_2, and p_3 as functions of the other variables and obtain a single expression:

$$dp_o/dt = G(p_+, p_-, dp_+/dt, dp_-/dt, p_o), \quad (12)$$

where the function G will depend on the functions f_c and f_e adopted for the conditioning and extinction functions.

Now, if the curves representing the values of p_+ and p_- are determined experimentally, we can express these variables as analytic functions of time. We can also obtain expressions for dp_+/dt and dp_-/dt as functions of time. Substituting the functions for p_+,

p_-, dp_+/dt and dp_-/dt in equation 12 we obtain:

$$dp_o/dt = G'(t, p_o). \quad (13)$$

If this differential equation can be solved we obtain:

$$p_o = f_o(t). \quad (14)$$

This equation will give us the desired value of p_o for any point in time during the experiment.

SUMMARY

In many discrimination learning situations some response, such as an orienting response, will be required of S before he is exposed to the discriminative stimuli. We call these responses "observing responses" (R_o), and indicate their probability of occurrence as p_o. Increases in p_o will result in increased exposure to the discriminative stimuli, and hence increased opportunity for S to learn or manifest discrimination. Decreased p_o will have the opposite effect. These results are operationally equivalent to decreases or increases in stimulus generalization between the discriminative stimuli. The following general hypothesis regarding changes in p_o can be derived from the principle of secondary reinforcement.

Hypothesis: Exposure to discriminative stimuli will have a reinforcing effect on the observing response to the extent that S has learned to respond differently to the two discriminative stimuli.

From this general hypothesis we derive the following specific hypotheses:

1. p_o will increase (or remain high) under conditions of differential reinforcement (discrimination training);
2. p_o will decrease (or remain low) under conditions of nondifferential reinforcement;
3. when a well established discrimination is reversed, p_o will decrease temporarily and then recover;
4. If the degree of discrimination and p_o are both low, the formation of discrimination will be retarded for some interval but will finally occur quite rapidly.

Evidence in support of these specific hypotheses was obtained in an experiment in which an R_o was measured directly.

This formulation may be useful for interpreting behavior in cases where changes in generalization between stimuli occur, and where the ease of formation of discrimination on the basis of some particular set of stimuli, changes as a function of training. *Ss* learn reversed discriminations more and more rapidly if reversals are presented repeatedly. The present formulation offers a relatively simple and readily testable interpretation of this phenomenon.

This formulation lends itself to precise quantitative statement. A quantitative analysis could be used in two ways: (1) to make quantitative predictions of behavior based on some set of theoretical statements regarding the component learning processes, and (2) to evaluate p_o from observations of measurable aspects of effective responses. The steps required for such an analysis are outlined.

REFERENCES

1. BERLYNE, D. E. Attention, perception and behavior theory. PSYCHOL. REV., 1951, **58**, 137–146.
2. BITTERMAN, M. E., & COATE, W. B. Some new experiments on the nature of discrimination learning in the rat. *J. comp. Psychol.*, 1950, **43**, 198–210.
3. BUSH, R. R., & MOSTELLER, F. A mathematical model for learning. PSYCHOL. REV., 1951, **58**, 313–323.
4. DINSMOOR, J. A. A quantitative comparison of the discriminative and reinforc-

ing functions of a stimulus. *J. exp. Psychol.*, 1950, **40**, 458–472.

5. EHRENFREUND, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. Psychol.*, 1948, **41**, 408–422.
6. ESTES, W. K. Toward a statistical theory of learning. PSYCHOL. REV., 1950, **57**, 94–107.
7. FRICK, F. C. An analysis of an operant discrimination. *J. Psychol.*, 1948, **26**, 93–123.
8. HARLOW, H. F. Studies in discrimination learning by monkeys: I. The learning of discrimination series and the reversal of a discrimination series. *J. gen. Psychol.*, 1944, **30**, 3–12.
9. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century Co., Inc., 1943.
10. ——. Simple qualitative discrimination learning. PSYCHOL. REV., 1950, **57**, 303–313.
11. KRECHEVSKY, I. "Hypotheses" in rats. PSYCHOL. REV., 1932, **39**, 516–532.
12. ——. A study of the continuity of the problem-solving process. PSYCHOL. REV., 1938, **45**, 107–133.
13. MCCULLOCH, T. L., & PRATT, J. C. A study of the pre-solution period in weight discrimination by white rats. *J. comp. Psychol.*, 1934, **18**, 271–290.
14. NORTH, A. J. Improvement in successive discrimination reversals. *J. comp. Psychol.*, 1950, **43**, 422–460.
15. ——. Performance during an extended series of discrimination reversals. *J. comp. Psychol.*, 1950, **43**, 461–470.
16. NOTTERMAN, J. M. A study of some relations among aperiodic reinforcement, discrimination training, and secondary reinforcement. *J. exp. Psychol.*, 1951, **41**, 161–169.
17. SCHOENFELD, W. N., ANTONITIS, J. J., & BERSH, P. J. A preliminary study of training conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1950, **40**, 40–45.
18. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
19. SPENCE, K. W. The nature of discrimination learning in animals. PSYCHOL. REV., 1936, **43**, 427–449.
20. ——. Continuous versus non-continuous interpretations of discrimination learning. PSYCHOL. REV., 1940, **47**, 271–288.
21. WILCOXON, H. C., HAYS, RUTH, & HULL, C. L. A preliminary determination of the function relationship of effective reaction potential ($_{s}\bar{E}_{R}$) to the ordinal number of vincentized extinction reactions ($\bar{n}$). *J. exp. Psychol.*, 1950, **40**, 194–199.
22. WYCKOFF, L. B. The role of observing responses in discrimination learning: Part II. Unpublished Ph.D. thesis, Indiana Univ., 1951.

[MS. received September 19, 1951]

THE MEASUREMENT OF DISCRIMINATIVE BEHAVIOR [1]

BY LEE S. CHRISTIE

Research Laboratory of Electronics,
Massachusetts Institute of Technology

Psychologists make experiments in order to record observations of behavior. Not every observable feature of behavior is recorded, but a selection is made which is determined by the purpose of the experiment. The experimenter is concerned not only with observations but also with some theory or theories of behavior. His observations have meaning in terms of, and *only* in terms of, theory. The purpose of an experiment is to gather meaningful data.

A theory of behavior is a structure which specifies relations between observable data and some set of hypothetical entities which, whether they are conceived of as observable or unobservable, are, in fact, unobserved. A theory is written in terms of variables, and its applicability depends upon the possibility of making inferences from the data of observation to the values of the variables. When the relation between data and theoretical variables is loose or ambiguous we call the data an indicator for the variable; when the relation is strict and unambiguous we say the data are a measure of the variable.

The development of measures is a part of theory. The selection and therefore the very collection of data is dependent upon measures or indicators. It follows, then, that it is not in general possible to collect large bodies of data at random and then construct a reasonable theory. The data that are collected depend upon the theoretical reasons for which they are collected. The development of *appropriate* measures and of *adequate* theory must go hand in hand.

Almost without exception psychological experiments have resulted in data which bear the indicator relation to the variables of the theories they were designed to test. Of necessity this has been the case, so no condemnation is in order. But the task of providing measures remains and becomes more and more pressing as psychological theory grows in scope and improves in definition.

Psychologists have employed many different types of data as bases for assertions about what is theoretically the same variable. There is seldom reason to question the relevance of the data to the theoretical question at hand, yet it is a fact that different proposed measures of the "same" variable often lead to conclusions at variance with one another. Since material of this conflicting sort exists, one way to clarify the situation and to move toward the development of true measures is to analyze the logical character of the indicators which have been used and bring into clear view the relations among them.

A particular problem which exemplifies the above situation arises in the theory and measurement of discriminative behavior. If discriminative behavior is observed in two different situations, the two sets of observations cannot be said to measure, in any satisfactory way, the learning that occurs, unless a logical basis for comparing the measures can be provided. When discrimination is studied with paired

[1] The work in this paper was done while the writer held an appointment as University Fellow in Psychology at the University of California at Los Angeles.

presentation of the stimuli and at another time with successive single presentation of the stimuli, usually different types of observation are taken. There seems to be no way to obtain comparable measures of the strengths of the response tendencies to the two stimuli across different methods of presentation.

We shall show how this impasse can be resolved. This paper will analyze measures of discrimination learning by using a psychological interpretation of a formal logical structure. By means of this technique of logical analysis, important conclusions can be derived concerning certain aspects of discriminative behavior and a solution can be given for an important instance of the problem of the comparability of measures obtained in differing situations. This result will be an important step toward the achievement of true measures of learning.

The tool we shall employ is a logical form of the theory of probability which permits both mathematical and logical operations within the same calculus (**2**). The psychological interpretation of this probability calculus identifies probability as the limit of the relative frequency of response to opportunity for response. With this definition it is obvious that the most commonly reported measure of behavior in discrimination problems, proportion of trials correct, is a relative frequency and hence an empirical estimate of a probability. An experimenter records this relative frequency because he takes it as an indicator of the strength of a response tendency. It should be noted that employing this procedure is independent of current theoretical predelictions; that is to say, it is done whether the psychologist calls the underlying variable a readiness, a reflex strength, a reaction potential, or some other construct. It should be noted further that although measures like proportion of trials correct are very commonly used, their theoretical, and especially their logical foundations, have not been analyzed as such. It is to this problem that the present paper is directed.

THE THEORY OF BEHAVIORAL PROBABILITIES

The assumptions and deductions which comprise the elements of our theory of behavioral probabilities follow. First, however, we give the definitions of the symbols which will be needed for the development of the theory.

$A =_{Df}$ the situation in which the animal is placed to start each trial.

$B =_{Df}$ any responses by the animal which leave it in the situation A (including the lack of any observable response).

$\bar{B} \equiv$ the approach response to the discriminative stimulus which carries the animal out of the situation A (read $\bar{B}$ as "not B").

$P(A,B) \equiv$ the probability that, starting from the situation A, the response B occurs.

The symbols A and B designate classes of events, and the symbol $\bar{B}$ designates the complement class of class B. The symbol P is a functor taking classes of events as arguments and having a numerical value. Thus, the expression $P(A,B) = p$ says: the probability of the behavior B in the situation A has the numerical value p.

In formulating the necessary probability statements we shall need the

following further symbolism: A dot, ., between two class symbols indicates the logical product of the classes, i.e., $A.B$ includes those events which are included both in A and in B. A hook, $\vee$, between two class symbols indicates the logical sum of the classes, i.e., $A \vee B$ includes those events which are included either in A or B, and possibly, but not necessarily, included both in A and in B. We sometimes need, also, a means of indicating the time point at which the psychological situation of an organism is asserted to be an event of a given class. For this purpose we shall superscript the event class symbol with a symbol giving the temporal designation. Thus an expression meaning, "If an organism is in the situation A and is making only responses of class B (ones which will not carry it out of the situation A) at time $t = 0$, the probability that it will be making a response also of the class B at a time infinitesimally later ($t = dt$), is equal to the value p" would be written:

$$P(A.B,B^{dt}) = p.$$

In such expressions we shall always omit the time designation 0, i.e., for B^0 we shall write simply B.

First Assumption

If a probability of response to a stimulus exists, the numerical value of the probability that the response will occur within a short interval of time is proportional to the size of the time interval. This proportionality is constant in time so long as the stimulus situation remains as initially defined,[2] and no approach to the discriminative stimulus is emitted. In symbolic form this assumption is written:

$$P(A.B,\bar{B}^{dt}) = \alpha dt.$$

This expression is to be read, "If A and B hold at some time t, then the probability that $\bar{B}$ holds at time $t + dt$, is equal to αdt, where α is a constant."

This assumption formulates the structure of the event sequence governed by the probability $P(A,B)$ as an instance of a continuous probability sequence, and as such, also as a sequence with probability transfer (2). The constant α is called the *transfer coefficient* of a continuous probability sequence and it measures the strength of the tendency of the sequence to switch from B to $\bar{B}$. This type of organization is the probability model for causal chains. Essentially, then, this first assumption asserts that the states of an organism considered in temporal order constitute a causal chain. More precisely, we postulate that the probability of response at any instant depends only upon the state of the organism at the immediately preceding instant, without reference to how that state may have been brought about by events which, in turn, preceded it.

Second Assumption

The probability that no response occurs within the time interval t, where the interval t is measured from a zero point at which the response of the organism is B, is a continuous and differentiable function of t. Since

[2] The stimulus situation may be defined with various degrees of restriction. We may consider the situation defined so that it terminates only when the organism makes the final approach response to the discriminative stimulus. This is the approach taken here. We may, on the other hand, break this period up into parts. We consider a series of situations, each of which ends when the orientation of the animal changes from one aspect of the total stimulus situation to another. This is the VTE problem and it will be treated in a later paper.

this postulate requires a probability to be a function of time, we shall write the probability as $f(t)$. The symbolic definition of $f(t)$ is then

$$f(t) = \lim_{\substack{dt \to 0 \\ ndt = t}} P(A.B,B^{dt}.B^{2dt}.\cdots.B^{ndt}).$$

The meaning of the limit symbol,

$$\lim_{\substack{dt \to 0, \\ ndt = t}}$$

is that the time interval t is divided up into more and more parts, each of which gets smaller and smaller but always subject to the restriction that the sum of the parts, *viz.*, ndt, is equal to the whole interval, t. This assumption is that the function, $f(t)$ is continuous and differentiable in t.

Third Assumption

The sequence of response instants starts with a B. Symbolically, it is expressed $f(0) = 1$; in words, only sequences starting with B are considered, and for such sequences and only for such sequences is it the case that the probability of B at the time $t = 0$ is equal to 1.

This set of assumptions has a very broad generality. They have been taken as our starting point, in the present instance, in order to give a theory of the measures of discriminative behavior, but they are equally applicable as a starting point for the theoretical consideration of any sort of behavior whatsoever. From the standpoint of the modern theory of causality they can be shown to be involved in any attempt to treat behavior as causally understandable. The first assumption merely lays down the principle that a behavior is causally determined by its immediate organic antecedents. The second assumption is needed primarily to expedite the mathematical treatment and its psychological content is unimportant. The third assumption concerns initial conditions in a way which is always satisfied in practice, that is, predictions of behavior are made from a knowledge of the antecedent conditions of the behavior.

We now proceed with the results of deductions in the general theory which will give 1) an identification of α, the transfer coefficient, as an exponential constant in a probability expression; 2) an explicit expression for the time course of the probability of response; 3) the distribution function for latencies of response. We can derive[3] from the assumptions the following relation

$$f(t) = e^{-\alpha t}. \qquad (1)$$

This result justifies the reference to α as the exponential constant, and gives the explicit expression for $f(t)$.

The meaning of (1) will be clearer if we consider it applied to the distribution of response times of a group of animals having identical response tendencies and placed in identical test situations. For each animal the probability of response within any short time interval is the same as for every other animal, but as time goes on more and more animals have responded and, hence, the number remaining from which responses with longer latencies can come is reduced more and more. In short, the number responding per unit time decreases exponentially. The structure is analogous to the physical example of radioactive decay. The case of a single animal introduced repeatedly into the test situation is explained similarly, since the chance of a response occurring with any given latency is run through on a particular trial only if the given latency is shorter than the actual latency shown on the trial. Thus, the occurrence of

[3] See appendix: Derivation A.

a response stops the sequence. There is, then, no possibility of a longer latency on that trial.

We shall use the symbol $\hat{f}(t)$ for the probability that a response occurs at least once in a time interval, t. For this probability function we obtain;

$$\hat{f}(t) = 1 - e^{-\alpha t}. \qquad (2)$$

The distribution function for latencies will be given by the product of the right hand sides of (1) above and assumption 1, since a latency is a time interval during which the response status of the organism is continuously B, and which is terminated at the instant that $\bar{B}$ occurs.[4] Using $g(t)dt$ for this distribution function, we find;

$$g(t)dt = e^{-\alpha t}\alpha dt. \qquad (3)$$

That this is a highly skewed distribution function conforms to expectation from common observation of latencies. Subsequent to the development of the ideas in this paper an article by Mueller appeared which reports data giving strong support to the special form of distribution derived from our postulates (1).

THEORY OF DISCRIMINATION LEARNING MEASURES

We proceed, now, to apply the results of the foregoing derivations to behavior in discrimination learning problems. As we stated earlier, the usual measure of learning of a discrimination is proportion of trials correct. This measure is a statistical estimate of a probability. We must deduce how such a behavior probability, e. g., Tolman's behavior ratio, is related to the basic probability of response.

[4] The theory of the distribution of latencies in situations other than the simple one of single presentation of stimuli is more complicated. It will be developed in subsequent papers.

Let us employ C and W for, respectively, proportion of responses correct and proportion of responses wrong, and p and q for, respectively, probability of response to rewarded (correct) and non-rewarded (wrong) stimuli. Let B_1 represent the correct response (response to the rewarded stimulus) and B_2 the wrong response (response to the non-rewarded stimulus). If $B_2 \equiv \bar{B}_1$, that is, if what we call response B_2 is simply the failure to make response B_1, then B_1 and B_2 exhaust the behavior possibilities of the situation. As a result, for this case we have, $C = p$, and $W = q$. Because these relations hold, and because it is always the case that, $C + W = 1$, we have $p + q = 1$. This case is a degenerate one in which there is no difference in value between C and p, nor between W and q. In psychological experimentation this structure corresponds to the situation on each trial in the classical conditioning paradigm. We shall say that the organism has a "real" choice when, as in a typical discrimination problem trial, the organism has the possibility of responding to either of two stimuli or may fail to respond to either. The meanings of p and q are unchanged, but C and W are different. When "trial" is specified to include responses to either of two stimuli but to exclude cases of failure to respond, we mean by C, proportion of trials correct, the relative frequency of responses to the correct stimulus counted among trials so defined. The value C is then a statistical estimate of the probability that if the subject is in the test situation and responds with either response B_1 or response B_2, then the response made is B_1. Symbolically,

$$C = P(A[B_1 \vee B_2],B_1).$$

By the rules of the calculus of prob-

ability we obtain from this expression,

$$C = \frac{P(A,B_1.[B_1 \vee B_2])}{P(A,[B_1 \vee B_2])} = \frac{P(A,B_1)}{P(A,B_1) + P(A,B_2)}.$$

The probabilities in this expression are our basic probabilities of response, p and q, so, finally we have,

$$C = \frac{p}{p+q} \quad \text{and correspondingly,}$$

$$W = \frac{q}{p+q}. \quad (4)$$

The above relations are valid only when each trial provides merely the choice between responding to a single presented stimulus or failing to respond and also when there are two kinds of trials, *viz.*, those in which the correct stimulus is presented and those in which the wrong stimulus is presented. This is the experimental paradigm we call "single stimulus presentation." The alternative, and more common, way of presenting the stimuli, i.e., both stimuli presented on each trial, we call "paired stimulus presentation."

We must develop the theory for this usual case of simultaneous (paired) presentation. We shall restrict ourselves to the case in which the processes of emission of approach responses to the two stimuli are independent. This is a significant restriction. Because of it we cannot expect the theory developed here for paired stimulus presentation to apply when the choice point behavior is characterized by VTE. However, in many discrimination problems, particularly after the discrimination has been fairly well learned, VTE's are infrequent. In this case the probability structure is an instance of what is known as the *competition of chances* (2). In this arrangement the probability of each of two events is reduced from what the probability would be if the chances of the two exclusive events were not operating concurrently. The reduction in probability is not produced by neural interaction. This possibility has been ruled out by our assumption of independence. It is, rather, purely and simply a consequence of the formal structure of the probabilities defined to describe the psychological situation. The effect can be ascribed to the fact that the occurrence of one of the two events cuts off part of the time interval during which the other event might otherwise have occurred. Through the use of the probability calculus we can derive the proper values for r, the probability of response to the positive (rewarded) stimulus reduced by the competition of chances, and of s, the corresponding probability of response to the negative (non-rewarded) stimulus.

The probability value, r, which we need is the probability that a response is made to the correct stimulus before it is made to the wrong stimulus. The following expression can be derived.[5]

$$r = \frac{\beta}{\beta+\gamma}(1 - e^{-(\beta+\gamma)t}). \quad (5)$$

Similarly for the probability, s,

$$s = \frac{\gamma}{\beta+\gamma}(1 - e^{-(\beta+\gamma)t}). \quad (6)$$

With these values determined, we may now calculate the proportion correct responses for the paired presentation case.

$$C = \frac{r}{r+s} = \frac{\beta}{\beta+\gamma}. \quad (7)$$

The tools are now at hand to demonstrate how an uncritical use of the

[5] See appendix: Derivation B.

measure, proportion of trials correct, can result in erroneous interpretations of learning phenomena. Let us consider the following experimental design. Two groups of animals are trained to discriminate the same two stimuli. One is trained by the usual paired presentation method, the other, by single successive presentations of the stimuli. The attempt is made to equate the level of learning in the two groups in terms of proportion of trials correct. Then the method of training is switched between the two groups and proportion of trials correct is again measured. We shall need identifying subscripts for the quantities appearing in our expressions. A preceding subscript will refer to the method of running the training trials, a following subscript, to the method of running test trials. Thus, ${}_iC_j$ means the proportion of trials correct for a group having training under procedure i, measured while running under procedure j. We shall use "2" for "paired" and "1" for "single," as subscripts. With symbols so defined we derive[6] the three inequalities: ${}_1C_1 > {}_2C_1$, ${}_1C_2 > {}_2C_2$ and ${}_1C_2 > {}_2C_1$. Paraphrasing this result in verbal form, we have: If two groups are trained on a discrimination problem to the same criterion score in terms of proportion of trials correct, one group by paired presentation and the other group by single presentation, the amount learned by the single presentation training group will be greater than the amount learned by the paired presentation training group. This is the case regardless of which method is used to test amount learned provided only that both groups are tested by the same method (${}_1C_1 > {}_2C_1$ and ${}_1C_2 > {}_2C_2$). Furthermore, if the running procedures are interchanged, the group changed from paired to single presentation will make more errors than the group changed from single to paired presentation (${}_1C_2 > {}_2C_1$).[7] Finally, this predicted outcome of the experiment does not result from a difference in amount learned ascribable to the difference in training procedure. *The reason for the prediction lies in the difference in the meaning of the measure, proportion of trials correct, when it is applied under the conditions of the two different training methods.*

There is some reason to believe that a difference in learning rate between the two training procedures exists. If such a difference does exist, it can be shown only by testing against a null hypothesis which takes account of the results developed here. That is to say, the appropriate null hypothesis does not specify a zero difference in the proportion of trials correct after a fixed number of trials unless the measurement is taken under the same running procedure for each group.

It should be noted that the predictions derived above follow from the very simple assumptions from which we started. The only additional principle needed is a minimal characteristic of the learning process in a discrimination problem, that is, that the response probability to the positive stimulus increases, or the response probability to the negative stimulus decreases, or both. We have not required a specification of numerical values or of the analytic form of the learning function.

The conclusions reached demonstrate that proportion of trials correct, taken in itself, does not constitute a proper measure of amount learned, because its meaning changes with the

[6] See appendix: Derivation C.

[7] This prediction will be shown to be confirmed by experimental data in a subsequent paper.

situation to which it is applied. Furthermore, probability of response is unsatisfactory since, being a function of time, it does not have a fixed value. However, probability of response as a function of time depends on the exponential decay constant, and this value does have a standard meaning independent of the arrangement of stimuli by trials. If the experimental situation in the single stimulus case were so structured that immediately following each response the animal were to return to the initial situation the exponential constant would be identified with rate of response. A similar but more complex relation can be derived for the paired presentation case. The proposal to use the exponential constant as a measure is therefore not unprecedented since Skinner (3) has proposed rate of responding as a measure and has used it extensively. Skinner's justification for the measure is almost wholly heuristic.[8] What has been contributed by the present paper is a theoretical analysis of the measure and a demonstration of its wide range of applicability.

SUMMARY

Starting from simple assumptions concerning the relation of the state of an organism to its probability of response, deductions have been made using the calculus of probability and applied to an analysis of the measures of behavior used in discrimination learning problems. The results demonstrate that the meaning of the measure, proportion correct trials, is dependent upon the type of experimental situation to which it is applied. The transfer coefficient of a continuous probability sequency is shown to be an appropriate measure of strength of response tendency.

APPENDIX

Derivation A.

Assumption 1 states;

$$P(A.B,\bar{B}^{dt}) = \alpha dt. \tag{1}$$

From (1) we derive;

$$P(A.B,B^{dt}) = 1 - \alpha dt. \tag{2}$$

From assumptions 1 and 2, specializing for $t = dt$, we have,

$$P(A.B,B^{dt}) = f(0 + dt). \tag{3}$$

Expanding the right side of (3) in a Taylor series and dropping second and higher order differentials, since we shall go to the limit, $dt = 0$,

$$P(A.B,B^{dt}) = f(0) + f'(0)dt \tag{4}$$

Since, by assumption 3, $f(0) = 1$, comparison of (2) and (4) leads to the conclusion

$$\alpha = -f'(0). \tag{5}$$

This result identifies the exponential constant, α, with the initial slope of $f(t)$. By definition, and as justified by assumption 2, we have;

$$f(t) = \lim_{\substack{dt \to 0 \\ ndt = t}} P(A.B,B^{dt}.B^{2dt}.\cdots.B^{ndt}). \tag{6}$$

The right hand side of (6) can be factored into the terms, $P(A.B,B^{dt})$, $P(A.B^{dt},B^{2dt})$, $P(A.B^{2dt},B^{3dt})$, etc., since by the definition of probability transfer, the probability at any instant is dependent upon only the state at the immediately preceding instant. Furthermore, since B specifies the same state regardless of the time super-

[8] The relations among probability of response, behavior rate, the transfer coefficient and latency in the Skinner box involve interesting problems beyond the scope of the present paper. The theory of measures in the Skinner box will be presented in a further paper.

script, each of the factors has the same value as every other. Hence, the probability expression after the limit sign in (6), substituting for each factors the value given by (2), becomes: $(1 - \alpha dt)^n$. If we let $\alpha dt = \frac{\alpha t}{n} = \frac{-1}{m}$, we have also $n = -mat$. Making this substitution, and noting that as dt approaches zero, m goes to infinity, we transform (6) into;

$$f(t) = \left[\lim_{m \to \infty} \left(1 = \frac{1}{m}\right)^m\right]^{-\alpha t}.$$

The expression inside the large brackets is the definition of the transcendental number e, therefore we have,

$$f(t) = e^{-\alpha t}.$$

Derivation B.

The probability value, r, which we need is the probability that a response is made to the correct stimulus before it is made to the wrong stimulus. Because of the transfer character of the probability sequence and assuming the two response emission processes independent we have,

$$P(A.B_1^{dt}.B_1^{2dt}.\cdots B_1^{(n-1)dt}.B_2^{dt}.B_2^{2dt}.\cdots B_2^{(n-1)dt},B_1^{ndt}) = \beta dt,$$
$$P(A.B_1^{dt}.B_1^{2dt}\cdots B_1^{(n-1)dt}.B_2^{dt}.B_2^{2dt}\cdots B_2^{(n-1)dt},B_2^{ndt}) = \gamma dt. \quad (8)$$

Where β and γ are, respectively, the transfer coefficients for probability of response to the positive stimulus and the negative stimulus presented singly. The probability that both responses are emitted in the same time interval, dt, is $\beta\gamma dt^2$. This is a second order differential and may be neglected. The case that the correct response is emitted in one of the intervals, $dt_1, dt_2, \cdots dt_n$ may be written,

$$\bar{B}_1^{dt} \vee \bar{B}_1^{2dt} \vee \cdots \vee \bar{B}_1^{ndt} \equiv \bar{B}_1^{dt} \vee B_1^{dt}.B_2^{dt}.\bar{B}_1^{2dt} \vee \cdots \vee B_1^{dt}\cdots$$
$$B_1^{(n-1)dt}.B_2^{dt}.\cdots B_2^{(n-1)dt}.\bar{B}_1^{ndt} \quad (9)$$

Since the terms on the right in (9) are exclusive, the corresponding probabilities may be added, therefore,

$$r = P(A,\bar{B}_1^{dt} \vee \bar{B}_1^{2dt} \vee \cdots \vee B_1^{ndt})$$
$$= P(A,\bar{B}_1^{dt}) + P(A,B_1^{dt}.B_2^{dt}.\bar{B}_1^{2dt}) + \cdots$$
$$+ P(A,B_1^{dt}.B_1^{2dt}\cdots B_1^{(n-1)dt}.B_2^{dt}.B_2^{2dt}\cdots B_2^{(n-1)dt}.\bar{B}_1^{ndt}). \quad (10)$$

Each of the terms on the right in (10) can be factored. Using relations (8) we obtain,

$$r = \beta dt + (1 - \beta dt)(1 - \gamma dt)\beta dt + \cdots + (1 - \beta dt)^{n-1}(1 - \gamma dt)^{n-1}\beta dt$$
$$= \beta dt \frac{(1 - \beta dt)^n(1 - \gamma dt)^n - 1}{(1 - \beta dt)\ (1 - \gamma dt)\ - 1}$$
$$= \frac{\beta}{\beta + \gamma - \beta\gamma dt}[1 - (1 - \beta dt)^n(1 - \gamma dt)^n]. \quad (11)$$

The term $\beta\ \gamma dt$ goes to zero as dt goes to zero and can hence be dropped. Putting $dt = \frac{t}{n}$ we have,

$$(1 - \beta dt)^n = \left(1 - \frac{\beta t}{n}\right)^n.$$

Using the definition of the transcendental number e, as before,

$$\lim_{n \to \infty} \left(1 - \frac{\beta t}{n}\right)^n = e^{-\beta t}.$$

Therefore (11) becomes

$$r = \frac{\beta}{\beta + \gamma}(1 - e^{-(\beta+\gamma)t}).$$

Similarly for the probability, s.

$$s = \frac{\gamma}{\beta + \gamma}(1 - e^{-(\beta+\gamma)t}).$$

Derivation C.

From relation (2),[9] substituting 'p' for $\hat{f}(t)$ and solving for α, we find:

$$\alpha = \frac{-1}{t}\log(1 - p).$$

Taking into account (4),[9] (5),[9] and (7),[9] we arrive at the four relations descriptive of the situation of our hypothetical experimental design:

$${}_1C_1 = \frac{p_1}{p_1 + q_1}, \tag{14}$$

$${}_1C_2 = \frac{\log(1 - p_1)}{\log(1 - p_1) + \log(1 - q_1)}, \tag{15}$$

$${}_2C_2 = \frac{\log(1 - p_2)}{\log(1 - p_2) + \log(1 - q_2)}, \tag{16}$$

$${}_2C_1 = \frac{p_2}{p_2 + q_2}. \tag{17}$$

The relation, $p > q$, will hold at the time of switching if discrimination learning is any process which increases the probability of response to the positive stimulus or decreases the probability of response to the negative stimulus, or both. It is readily seen that if $p > q$, then

$$\frac{|\log(1 - p)|}{p} > \frac{|\log(1 - q)|}{q}. \tag{18}$$

The limits on the values of p and q are 0 and 1, hence, both logarithms have negative signs. Noting this, we can modify (18) to the form;

$$\frac{p}{p + q} < \frac{\log(1 - p)}{\log(1 - p) + \log(1 - q)}. \tag{19}$$

Because of inequality (19) we infer that at the time of switching;

$${}_1C_2 > {}_1C_1 \quad \text{and} \quad {}_2C_2 > {}_2C_1. \tag{20}$$

But, by the conditions imposed by the experimental design, at this time;

$${}_2C_2 = {}_1C_1. \tag{21}$$

Therefore we conclude;

$${}_1C_1 > {}_2C_1, \tag{22}$$

$${}_1C_2 > {}_2C_2 \tag{23}$$

and

$${}_1C_2 > {}_2C_1. \tag{24}$$

[9] These relations are from main body of article.

REFERENCES

1. MUELLER, C. G. Theoretical relationships among some measures of conditioning. *Proc. nat. Acad. Sci., Wash.*, 1950, **36**, 123–130.
2. REICHENBACH, H. *Theory of probability*. Berkeley: Univ. California Press, 1949.
3. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.

[MS. received September 29, 1951]

THE BRAIN ANALOGY: A DISCUSSION [1]

BY H. EDGAR COBURN

Registered Civil Engineer, San Diego, California

Introduction

The Brain Analogy theory of behavior, presenting a point of view that differs in detail from some existing schools of thought, invites a re-appraisal of our concepts. The following discussion supplements the original paper (1).

Nature of the Brain Analogy

The Brain Analogy (BA) is one of several particular cases of the same general mechanism or concept. The significance of the relationship is that it seems to cast light on the nature and origin of intelligence. The general case, the Correlation Detector (CD), to give it a name, would be similar to the BA except for its symmetrical latency distribution and consequent indifference to the direction of conditioning.

The universe is rational, a fundamental premise of science, and therefore there is a fixed percentage of the S in any situation that are related to the US.

The S is always compound for complex animals, but initially this fact is of no significance to the nervous system because various types of neurons are collected into functional clusters (an inhibitory bipolar, for example, is functionally a mono-polar when the inhibitory pole is stimulated at a sub-threshold level). The components of the compound S act as one simple S of equivalent intensity, even though each component appropriates a share of the motor cells; however, this latter fact remains concealed from both the organism and experimenter as long as the program is unchanged—test trials omitted. For the purpose of analysis, then, we can assume the existence of elementary S.

Since the brain has no way of "knowing" which S are significant, it must necessarily *sample* the sensory universe and condition to all S on a probability basis. There is no *a priori* criterion for assigning probabilities, but these are determined by the nature of the environment, the nature of the organism, and the interaction of mutations and natural selection. Experimental observation indicates that magnitude and recency of S are of primary importance.

The simplest hypothesis concerning two elementary stimuli, CS and US, is that they are either related or random.

The unknown probability that the S are related is p, and that they are random is $q = 1 - p$. The probability of obtaining x related pairs in n sensory samples (or x stable, conditioned delta cells on n motor cells) is given by

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x},$$

which increases with a relative increase in CS over random S. This binomial distribution for discrete variables, approaches as a limit normal distribution with mean np and SD $= \sqrt{npq}$ as the number of samples increases (2, Chapter 3). The mean of the samples equals the "fixed percentage in a rational universe" and therefore the conditioning to any component will be approximately proportional to its relative strength *as defined in terms of the relative number of delta cells involved.* Since this is a fact that motor cells are appropriated by S components in proportion to their relative strengths (5, p. 142), we can infer that the probability of significance of S

[1] The author wishes to express his gratitude to Vera Jane Coburn for her assistance in preparing the manuscript.

is proportional to their relative magnitudes; differences between analyzers and between species are adjusted by mutations and natural selection which alter the relative number of delta cells.

The sampling, of course, is done only by the actuated and accessible motor cells. The stable, appropriated cells are not of sample distribution, except momentarily, because all but systematic connections are driven to extinction between formal trials and the remainder are therefore biased. After repeated trials, when the animal "knows" the correct signal, it nevertheless continues the process of random sampling by means of the accessible motor cells. If it were not thus, if "knowing" prevented further sampling, the organism would be at a serious disadvantage in all cases where its hypothesis was erroneous. Up to this point we have considered only an "instantaneous" sample, mostly ignoring the fact that S are also temporally distributed.

Curve CD (Fig. 1) implies that the probability of a relationship between two S is greatest when the S are closely related in time. But this is the general case, the CD, and is not a satisfactory hypothesis for an animal to react to because it fails to indicate which S comes first. A better hypothesis is curve CE, a qualitative distribution function representing cause and effect, which implies a particular temporal sequence—the CS followed by the US. For an abstract study of cause and effect, the CE curve may be satisfactory, but again we fail to establish a hypothesis suitable for animals because inertia has been ignored. It is at this point that the BA-type hypothesis enters the scene with provision for both temporal sequence and inertia.

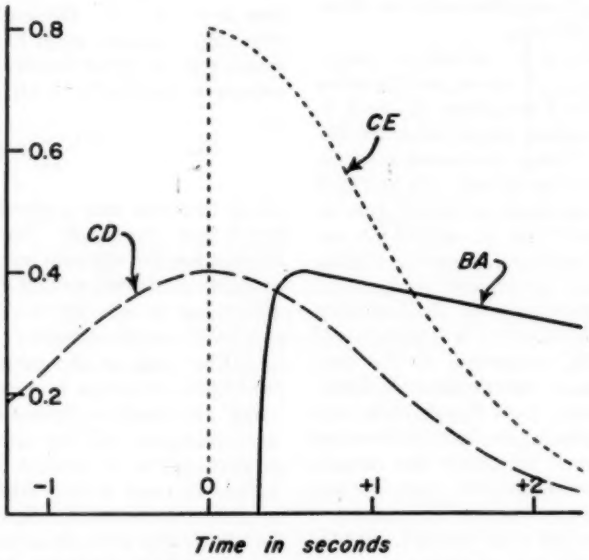

FIG. 1. The probability of relationship between CS and US is shown for each of three distributions. The area under each curve is unity (with logarithmic abscissas for BA). BA is the result of biologic evolution as explained in the text.

ORIGIN OF SKEWED LATENCY

Biological data suggest that organisms, and their nervous systems, vary due to unknown factors that influence the germ cells and that these hereditary variations or mutations, which apparently may be directed in more or less random evolutionary directions, are subject to an elimination process in the struggle for existence. As a consequence, nervous systems reflect the nature of the environment, and should therefore be in agreement with the CD-type hypothesis or one of the particular cases. The existence of inertia makes it impossible for an animal to respond correctly to signals which are too closely followed by the US; the failure is an obstacle to survival —the "correct" cerebral connection is futile. Therefore, the mutations which tend toward the general theoretical case, the CD, or even the CE, are eliminated in favor of the mutations which not only involve more latency for CS pulses than for US, but also delay CS pulses a minimum amount (1, Postulate 5) which should be roughly in proportion to the mass of the animal.

It is also worth noting that the BA complies with other evolutionary requirements to the extent of being a structure that can evolve by successive, but more or less imperceptible, steps or mutations *each one of which is functional.* In addition, it possesses the property of allowing extirpation of parts without rendering the entire mechanism inoperative. In fact, if the parts are located spatially in accordance with function, as appears to be true of organic nervous systems, comparable extirpations will have comparable results, and in general, the degree of loss in complexity of function will parallel the loss in parts. These considerations, important for learning theory, and vital in the selection of an adequate postulational system, will be made more evident in a proposed future paper on subcortical mechanisms.

VERBALIZATION

Intelligence has been defined as *the property of reacting on a basis of probability as determined by the individual organism's incomplete sensory samples of the environment* (1, p. 155).

No mechanism, organic or otherwise, has unlimited sensory possibilities. The limitations restrict the number of possible correlations that can be formed directly through the senses and without intervening variables. They restrict intelligence.

Although man has zero intelligence where X-rays are concerned (because he has no X-ray receptors), he can use instruments and verbalization to effect appropriate responses. But he is then showing intelligence for "pointer readings," say, and not for X-rays *per se.*

The introduction of intervening variables, instrument readings, printed data, etc., coupled with some form of verbalization process, increases the effective use of existing intelligence. However, an approach based on verbalization is unnecessarily redundant. In communication theory, too, verbalization has been found to be exceptionally inefficient; in fact, Shannon (8) and others have shown that verbal English occupies approximately 1000 times the frequency spectrum required to transmit the actual information involved.

The only excuse for the existence of verbalization is that we do not possess the appropriate receptors to make direct correlations. Fortunately, inorganic intelligence can avoid the handicap and for this reason, as well as for others, it can be much simpler for a given performance.

THE MEASURE OF INTELLIGENCE

The fact that all mammals are roughly equal in maze-learning ability (4, p.

109) seems to show that the conditioning mechanism itself is approximately the same and that there is an optimum rate of learning. The differences in intelligence, then, may be due essentially to the differences in analyzers. We refer, of course, not to the receptors themselves, but to the entire mechanism between *S* and delta cells.

Supporting evidence for the hypothesis is found in musical aptitude studies. Highly intelligent behavior, where pitch discrimination is concerned, is impossible if the auditory analyzer is below average. Nor will any amount of practice correct the difficulty (**7**, p. 58).

If differences in intelligence are due primarily to analyzer differences, as suggested by the above data, it is clear that the measure of intelligence is the ability to discriminate. Where an objective test is available, as with pitch discrimination, measurement is relatively simple. Where objective standards are difficult to establish, as with verbal discriminations, the results may be less satisfactory. With animals, other than man, a habit-saturation technique can be employed, but care must be used in the selection of a measure of the strength of the habit. The BA theory indicates that magnitude of response and resistance to extinction are inadequate. The true measure of habit strength is the percentage of motor cells, actuated during learning, appropriated by stable connections. Not only is there no means as yet of making this measurement, but the size of the stable group will fluctuate with changes in the experimental program. Furthermore, a complete appraisal of intelligence requires tests of all analyzers.

OPTIMUM LEARNING RATE

It is probable that the interaction of various mutations and natural selection has established a rate of conditioning which closely approximates the theoretical optimum. While slow learning is obviously detrimental, extremely fast learning may likewise be injurious when we recall that one or two sensory samples do not determine an answer but merely establish a hypothesis. *The brain has no way of "knowing" which elements of the stimulus situation are significant.* It must necessarily accept and condition to all *S*, automatically, and in accordance with probability. The *appearance* of purpose is due solely to the fact that systematic *S* result in *stable* conditioning. This is exactly analogous to the fact that organisms, in general, appear to be purposive due to extermination of "non-purposive" mutations in the struggle for existence. A sample of *stable* habits, like a sample of *living* species, is grossly biased.

Let us start by assuming that conditioning is complete in one trial. Although nearly all actuated motor cells are appropriated within approximately 0.3 second, roughly 90 per cent of the CS elements are random or at least unstable, and 10 per cent represent the conventional CS. The same random elements, of course, are not necessarily actuated each trial, some are actuated between trials with consequent extinction, and the stable elements gradually assume complete control.

Mutations and selection would increase the conditioning rate for the conventional CS *only if there were no disadvantage in a faster rate.* The fastest possible satisfactory rate is that which still maintains the elements of the sensory environment in the proportion of their probable significance, which is measured by amplitude, recency, etc. The proportional distribution of the stimuli insures that minority signal elements are not unduly discriminated against and hence the increment of the conventional CS is approximately 10 per cent (less for minority elements); otherwise, we have a deviation from the definition of intelligence.

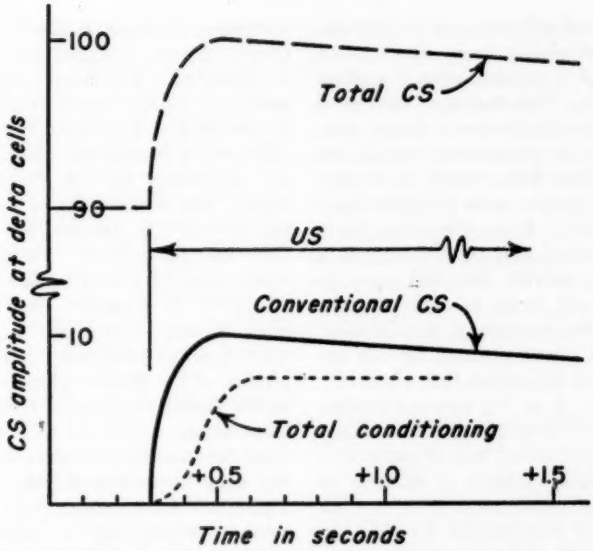

FIG. 2. The conventional CS is shown as a relatively small part of the total CS in conformity with the fact that conditioning is on a probability basis. While the conventional CS is entirely arbitrary, the brain, making no such distinction, automatically accepts all stimuli and conditions to them in proportion to their relative magnitudes at the delta cells. The total-conditioning curve (on a separate vertical scale) flattens off at 100 per cent appropriation of motor cells within approximately 0.3 second after beginning of US. In this figure US is timed to start at the end of minimum delta cell latency and the overt conventional CS is an auditory "click."

The *effective duration* is now seen to be the temporal distribution of conditioning of a population of delta cells, rather than an independent postulate, and it is unnecessary for the motor cell to disable itself (1, Postulate 13). Continuing in a qualitative vein concerning the nature of the temporal distribution function, and assuming SD = 0.075 second, enough motor cells are still available for a longer time than minimum sensory cell latency (of the conventional CS) to permit a small percentage of backward conditioning. This is transitory (5, p. 393), except for interference by gamma phase differentiation,[2] because of displacement by less latent elements of random *S*. The possible backward conditioning occurs, of course, only when the experimental program is appropriate.

It is evident that isolated action of a conventional CS is without foundation in fact. The actuated motor group is always quickly appropriated by delta cells because zero sensory stimulation is nonexistent and weak stimulation is largely offset by strong *Q* (1, Postulate 12). The conventional CS is merely part of the ever-present, relatively large, sensory universe (see Fig. 2).

REINFORCEMENT

The original BA paper specified that reinforcement of conditioned and unconditioned delta cells are separate phenomena (1, p. 161). The mechanism of reinforcement for unconditioned delta cells is unsatisfactory for reinforcement

[2] Formerly called *periodic reconditioning* (1, p. 162).

of conditioned cells because the former, having a probability basis, would cause instability of S differentiation if applied to the latter. This fact was discovered too late to be dignified as a formal postulate. And no explanatory mechanism is being offered here, except to observe the implication that some additional factor is involved. Along this same line it should be noted that conditioning of a delta cell by another delta cell pulse, in the motor cell body, is expressly forbidden by the presence of the disabling function. Current research by the author supports the belief that secondary conditioning (5, p. 33) does not involve such delta cell-to-delta cell conditioning.

The basic cause of lack of agreement between various schools of thought on the nature of reinforcement, stems from the fact that instrumental conditioning and allied phenomena are inherently much more complex than the idealized salivary reflex. It is equally impossible to give convincing support to the BA concept of reinforcement, which embraces both classical and instrumental conditioning, without employing properties of subcortical mechanisms. It is easier, but far from satisfying, to show that some of the accepted views on reinforcement are seriously compromised by available evidence.

Food reinforcement for the conditioned salivary reflex may be quite different from feeding at the end of a maze. The latter, usually considered to be reinforcement, in spite of latent learning experiments, terminates the drive (at least temporarily) and thereby causes a selective action to take place. The possible UR's are elicited during drive by the various random US (reinforcement) in the presence of the CS (1, p. 177). The order being somewhat random, due partly to fluctuating thresholds, indicates that the CR's which form will not all receive reinforcement each trial. The exception, of course, is the "correct" response because each trial continues until it is elicited. The correct habit, never receiving non-reinforced trials, ascends to the head of the hierarchy; the other CR's lead a fluctuating existence but do not necessarily become completely extinct. The termination of drive, after the correct CR, assures that all other CR's will be insecure. Failure to terminate the drive would allow the effect of the CS to continue, after reinforcement by the appropriate random US, with resulting decrement following depletion of the gamma phase. In other words, *habits continually form and decline during drive, but if a systematic drive-termination procedure is followed, the habit formation becomes selective.* This gives the appearance of reinforcement by feeding, say, but this "need reduction" is not only incidental as far as conditioning is concerned, it actually *follows* conditioning. The advantage of this hypothesis is that it applies equally well to glandular reflexes being of the S–R type.

Another example concerns the avoidance of shock. It has been said that the avoidance of shock—the absence of stimulation—is reinforcement, yet it has been shown that the animal learns even though shock is never avoided (3, p. 7).

Some of the difficulties concerning the nature of reinforcement disappear if we drop verbalization, which conceals inconsistencies, and substitute physical structures which expose them. It isn't easy to look at a physical nervous system, biologic or analogic, even on paper, and imagine that absence and presence of stimulation are identical. In a physical structure the concept of reinforcement has to be "pinned down." The BA probably contains as many inconsistencies as some other systems but part of its strength lies in the fact that they are visible rather than concealed.

MOLAR VS. MOLECULAR THEORIES

The failure of "physiological" learning theories to give explicit answers to numerous problems has, in the past decade, given rise to suggestions to abandon the molecular approach. Reliance is to be placed on molar theories, ignoring the actual mechanism involved, and substituting more or less abstract constructs. It has been said that, after all, the psychologist is concerned with behavior, and molar constructs offer a practical way out of the difficulties of neurological explanations. No doubt, the escape clause is attractive, provided the underlying assumption is valid.

The author worked along molar lines by using hypothetical nerves rather than neurons until the discovery of the *principle* involved in *S* differentiation—exclusive appropriation of motor cells by means of the disabling function (1, Postulate 13)—led to a consideration of the required *mechanism*.

It is apparent, of course, that nerve properties are statistical properties of groups of neurons; it is not clearly evident, however, that nerve properties possess serious limitations for learning theory. Stimulus differentiation, for example, *can* be accounted for by nerve properties, provided no change is made during the experimental program, but the restriction can be removed only by assigning the explanation to neurons because altering the program, in effect, calls for splitting the nerves into new stable and unstable groups (1, p. 169), thereby destroying their identities from a molar point of view. A major restriction, operating at such a primitive level as *S* differentiation, is unlikely to be tolerated for long in psychological *theory*.

The alternate is to dig deeper into the mechanism, even though the complexities appear to be multiplied, and consider the elementary properties of neurons; not only will these elementary properties provide answers to many phenomena which would require unnecessary elaborations at the molar level, but being elementary, are better able to survive the important test of interchangeable constants. In particular, it then becomes feasible to base theory on a physically-realizable structure and to comply with the dictum of physics that *all constructs must be definable and detectable by means of experiments that are at least conceptually possible* (6, p. 78).

CONCLUSION

The foregoing discussion must necessarily be considered heuristic, and, as always with science, provisional. It seems reasonable to conclude, however, that the importance of evolution for behavior theory has been demonstrated. The author would also like to believe that the nature of reinforcement, probably the outstanding problem in psychological theory, has been clarified a little.

While the implied "normal" distribution of the significance of the elements of the sensory universe is clearly of intuitive origin, and the top-heavy predominance of "random" conditioning, an apparently necessary corollary, seems extreme, it is offered because the most acceptable measure of the value of a hypothesis is usefulness: It seems to work.

REFERENCES

1. COBURN, H. E. The brain analogy. PSYCHOL. REV., 1951, **58**, 155–178.
2. HOEL, P. G. *Introduction to mathematical statistics*. New York: John Wiley & Sons, Inc., 1947.
3. LIDDELL, H. S., JAMES, W. T., & ANDERSON, O. D. The comparative physiology of the conditioned motor reflex. *Comp. Psychol. Monogr.*, 1934, **11**, No. 1.
4. MORGAN, C. T. *Physiological psychology*. New York: McGraw-Hill Book Company, Inc., 1943.

5. PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford University Press, 1927.

6. PERSICO, E. *Fundamentals of quantum mechanics.* (Trans. by G. M. Temmer.) New York: Prentice-Hall, Inc., 1950.

7. SEASHORE, C. E. *Psychology of music.* New York: McGraw-Hill Book Company, Inc., 1938.

8. SHANNON, C. E. Recent developments in communication theory. *Electronics,* 1950, **23**, No. 4 (April), 83.

[MS. received October 12, 1951]

A "COMMUNICATION ANALYSIS" OF CONCEPT LEARNING

BY CARL I. HOVLAND

Yale University [1]

In concept-formation experiments the items presented in the learning series may be of two different types: "positive instances," which are examples of the concept and include the essential characteristics, or "negative instances," which lack one or more of the necessary characteristics, and are therefore examples of what the concept "is not." The contribution of each type of item to the learning of concepts has been extensively studied by Smoke (5, 6). He found that negative instances are of little aid in learning the correct concept. What is not clear from his study and from similar investigations, however, is whether the ineffectiveness of negative instances is primarily attributable to their low value as carriers of information, or whether it is primarily due to the difficulty of assimilating the information which they do convey.

The foregoing statement of the alternatives indicates that the concept-formation experiment may be regarded as a communication situation in which *E* transmits the combination of elements he has selected as constituting the concept through a series of messages, some of which are labelled "correct" (positive instances) and others "incorrect" (negative instances). This formulation stresses the importance of first analyzing the information transmitted by positive and negative instances, and then studying experimentally the relative difficulty of assimilating material presented via positive as compared with negative instances, equated with respect to the amount of information conveyed. The present paper is devoted to the first task, of analyzing from a theoretical standpoint the amount of information required to specify a concept by positive or by negative instances.

The difficulty with the usual concept-formation set-up for the present problem is its unclear structure. In the traditional concept-formation experiment, a whole series of stimulus figures containing combinations of characteristics constituting the concepts are presented to *S* together with paired names or nonsense syllables. *S*s are required to give associated names at the sight of the appropriate stimulus figures. In Smoke's experiment, for example, the syllable "dax" was to be associated with the concept of a circle with one dot inside and one dot outside. Variations were employed from instance to instance in the size and color of circle and the exact position of the dots. Each positive instance was marked with a plus sign to indicate that it was an example of the class in question. Negative instances were introduced with a minus sign to indicate that they did not conform to the definition of the concept.

What is quite unclear from such a method is *S*'s expectations as to the type of concept model to be employed and as to the number of characteristics of the stimulus figure which might be relevant to the concept-formation task. Without

[1] The writer wishes to express his appreciation to Walter Weiss, for suggestions made during discussions preparatory to experimental work, and for assistance in determining empirically some of the information presented in Table 1, George A. Miller for helpful suggestions and criticisms, particularly concerning notation, B. J. Underwood for careful reading of the manuscript, and Shizuo Kakutani for advice concerning one of the formulae.

some common understanding on the part of *E* and *S*, it is difficult to define what information is conveyed by each instance. Otherwise, factors not intended by *E* to be relevant may be selected as possibilities by *S*, and the information conveyed may thus not properly eliminate some of these. *S*s who are most imaginative in considering different possibilities may be penalized. To take an extreme example, let us imagine an *S* who considers the color of the card on which the stimulus figures are being presented to be a possible dimension to consider. He may on the basis of positive instances derive the concept that a "dax" is a circle with one dot inside and one dot outside *on a white background*. The extra dimension would be considered incorrect by *E*, but it would be logically correct on the basis of the data presented to *S*. There is a suggestion in Smoke's account that a result like this sometimes occurs. He points out that *S*s are frequently able to distinguish between correct and incorrect instances in the test series (prepared in terms of E's definition of relevant factors), but are unable to define the concept correctly. Their definitions are customarily described as being "too general." This might sometimes be due to *S*'s inclusion of characteristics which had been considered irrelevant by *E* and hence not systematically eliminated as hypotheses by the series of instances used.

There are, of course, many situations like Smoke's, where the concept-formation task is to determine the structure of the concept and ascertain which characteristics are relevant in an instructured and undefined situation. But because of the difficulties just mentioned, for the analysis of the present problem it will be clearer if we confine our attention to the type of concept-learning paradigm used by Oseas and Underwood (3). *S*s were presented with Weigl-type stimulus figures picturing squares, circles or triangles. Each figure could be one of three sizes (¼ inch, ½ inch, or ¾ inch) and one of three shades (black, gray, or white). The concept to be learned consisted of a combination of one value of one dimension (e.g., the *small* (¼ inch) value of the size dimension) with one value of another dimension (e.g., the *circle* from the shape category). This would give, as an example, a *small circle* as the concept. For any single concept the third factor was irrelevant (in this example, shade).

In the Oseas and Underwood study *S*s learned a whole series of different concepts and to each they associated an appropriate consonant letter. The role of positive and negative instances will be clearer if we restrict the analysis to the learning of a single concept. Let us also simplify the set-up further by considering first a task in which there are three dimensions involved (form, size and shade as above) but where each dimension has only two possible values (circle or square, large or small, black or white). *S* will be given complete knowledge of the set-up, i.e., he will know that these are the characteristics which will be used and that the correct concept will be a combination of one value on each of two dimensions [e.g., a *white square* (of any size) or a *large black* object (of any shape)]. *S*'s task thus consists of ascertaining which are the relevant dimensions and what values on each dimension are correct.

This type of model is schematized in Fig. 1. Each block represents an instance. The three characteristics of size, shade, and shape, will be referred to as *dimensions* and the different degrees of each characteristic as *values* on the dimension. In this figure there are eight possible instances. (Dimension A with 2 values × Dimension B with 2 values × Dimension C with 2 values.)

Let us first consider the most general case. If no information at all were

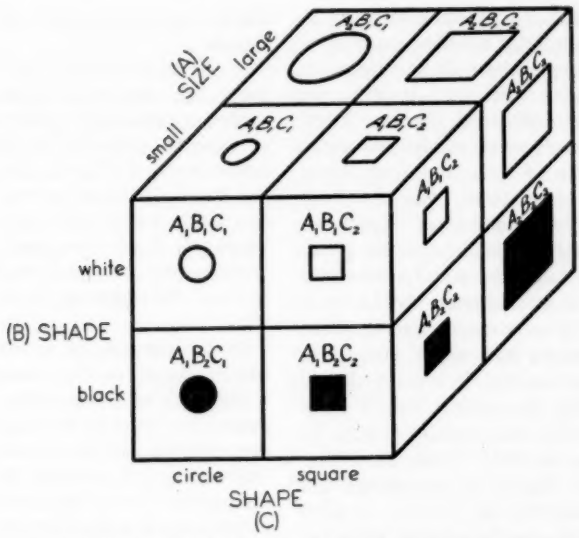

FIG. 1. Diagram showing total possible instances for three dimensions (size, shade, and shape) with *two* values of each.

given to *S* as to the nature of the concept, the dimensions involved, or the number of instances which are included, etc. he would have to consider 256 possibilities ("hypotheses") as to the combinations involved. These would include the one extreme case where all instances were to be included (where the concept was, for example, "geometrical figures") and the other where none belonged (e.g., "*diamond* shaped objects"). There could then be various definitions which would include single instances (8 possibilities), pairs (28), triplets (56), quadruplets (70), quintuplets (56), sextuplets (28), and septuplets (8). There are the familiar values of the binomial expansion and the total number of possibilities is the familiar $2^{(m^n)}$ involved in circuit analyses, with $m = 2$ for the number of values and $n = 3$ for the number of dimensions.

Under the conditions described every single one of the eight instances would have to be presented before the individual would be able to know what constituted the correct concept. And positive and negative instances would provide equivalent information, i.e., knowing that a particular instance was included and knowing that a particular instance was not included would be equally informative concerning the concept. This is the type of model involved in much of current information theory **(4)**.

If one were to tell *S* that all of the concepts would involve two out of the three dimensions (or one out of three) the number of hypotheses he would have to consider would be greatly reduced. He would then only have to consider the various combinations of dimensions times the number of possible arrangements among the instances involved within the two dimensions. In the case where two of the three dimensions were involved, there would then be three ways in which the dimensions could be chosen ($_3C_2$). The number of arrangements

within the dimensions would be 2^{2^2} or 16. The 16 possibilities would be made up of the one extreme where all four instances were included and the opposite one where all were excluded, plus four where single instances were included, six where combinations of two, and four where combinations of three, were involved. Thus the total number of hypotheses to be considered would be 48, the product of the combinations of dimensions (3) and the arrangements within (16).

Lastly, let us consider our situation where *S* knows how many dimensions there are to consider, how many will be necessary for the concept and in addition knows how many values of each dimension are required. Under these conditions the number of hypotheses will be determined by the number of ways the dimensions can be selected times the number of ways the values can be selected within the combination of dimensions. If *S* knows that two of the three dimensions will be relevant there will then be, as we have seen, three ways of picking combinations of two dimensions. Once these are chosen there will be only four ways $({}_2C_1)^2$ of choosing the instances in such a way that one value of each relevant dimension is involved. Thus there will be a total of only 12 hypotheses for *S* to consider (three for dimensions times four for values within the dimensions).

The remainder of the present paper will be devoted entirely to an analysis of the third type of model, where the concept is one involving a systematic arrangement of the instances along particular dimensions and with specified values of the relevant dimensions. *S* will be told what the characteristics of the concept are. Under these circumstances positive and negative instances may not convey equal amounts of information about the concept. Our problem is the relative effectiveness of the two kinds of instances under various circumstances.

If the third model is employed, how many instances would have to be presented to reduce the number of hypotheses from the original 12 to the single one defined as correct by *E*? Let us say that the concept chosen in our example is that of a *white circle* (of any size). There will then be two positive instances (*large white circle* and *small white circle*) and the remaining six will be negative.

In this example, if we use only positive instances in the concept learning series, both of the possible positive instances are required to transmit the correct concept. If the series is made up only of negative instances, one can reach the correct concept by successive elimination of each of the six wrong instances. Actually, however, only five of the possible six have to be transmitted since, as will be discussed below, one can deduce the correct concept with less than the total number of negative instances.

Armed with the knowledge of how many instances of each type are required, one can then pose one of the experimental questions studied by Hovland and Weiss (2) as to how accurately *S* will reach the correct concept when he is shown only the two positive instances as compared with the situation where he is shown only the five negative instances. The information provided is complete in either case, in the sense that a machine could be constructed which would reach the correct solution on the basis of the data presented. But the psychological processes of assimilating the two sets of information may be quite different, and one may be much more difficult than the other.[2]

[2] The two positive and five negative instances are of course only equivalent with respect to amount of information relevant to the definition of the concept. But the five negative instances convey a greater total

Analysis can be made of what possibilities are theoretically compatible with the conditions after varying numbers of each type of instance. After two negative instances, for example, how many possibilities remain? What must be analyzed is the number of "possible hypotheses" after each instance. With the example used above, there are 12 possible hypotheses as to the correct concept at the outset: *black circle* (of any size), *black square, white circle, white square, large circle* (of any color), *small circle, large square, small square, large black object* (of any shape), *small black, large white, small white.* After the first positive instance (e.g., a *small black circle*) *S* can narrow the alternatives to those in which the size and color given are correct and shape irrelevant; the size and shape are correct and color irrelevant; or the color and shape correct and size is irrelevant. In other words, he can know that the correct concept is one of the following: a *small circle* (of any color), a *black circle* (of any size) or a *small black* object (of any shape). Thus after the first positive instance he is left with *three* hypotheses. The second positive instance will narrow this down to *one* (the correct combination). Since this instance will be the same in two respects and different in the third he will see which two dimensions are essential and which can have either value and still be correct. Thus a second positive instance of a *small black square* will indicate that small and black are essential and that shape is irrelevant.

In the case of negative instances the individual also, of course, starts with 12 hypotheses. The first negative instance will eliminate three of these possible hypotheses. For example, being told that it is not a large black square rules out the possibility that the correct concept is a *large square,* a *black square,* or a *large black* object, since any combination where two of the dimensions were correct and the third irrelevant would make the instance correct. Three more possibilities can be removed with the second instance. For example, a negative instance of a small white circle rules out *small circle, white circle* and *small white object.* Thus he now has six of the original 12 hypotheses left. But the third negative instance necessarily involves some duplication of information since it confirms the exclusion of one of the previously eliminated hypotheses and eliminates only two additional hypotheses. (One describes this situation as "redundancy" in information theory.) Similarly, two more hypotheses are eliminated by the next instance, leaving two. The last instance will narrow the choice to the correct one, if all of the information previously presented is remembered correctly. Thus five instances will here eliminate 11 hypotheses, leaving only the correct one.

Sometimes the term "concept" is used for the case where the classification is based on a single characteristic or relationship (e.g., that of "circle" in Fig. 1). Under these conditions half of the eight instances would be positive and half negative. One would have to consider six possible hypotheses (each involving sets of four blocks in a single plane). It would take two positive instances [to specify the plane involved] or two negative instances [to rule out all but one possible plane]. In this type of situation the two types of instances are completely complementary and equally efficient carriers of information. This condition is similar to that typically considered in "information theory." Experimental data of Buss (1) suggests that under these conditions learning may be as efficient with negative as with positive cases.

---

amount of information than the two positives, even if the surplus information is irrelevant for the present task.

In the two alternative situations one is usually as interested in the excluded cases as the included. Thus knowing the bases for classifying some members of a species as male permits one to classify all others as female. But in the type of concept with which this paper is primarily concerned only the particular combination specified may be of interest and combinations having the excluded characteristics of little concern. For example one may be concerned with the concept "ruby" characterized as a red, translucent stone, and have no interest in the category of non-red, non-translucent stones.

Let us now consider the original Oseas and Underwood (3) example. They used three shapes, three sizes, and three shades. The concept was some combination of two of the three dimensions. Their concepts would then be described as having three dimensions (two relevant and one irrelevant) with three possible values of each. Only one of the three levels was correct for each of the two relevant dimensions. With this arrangement there are 27 possible instances ($3 \times 3 \times 3$) and of these *three* will be positive (e.g., *small black circle, small black square,* and *small black triangle*). There will be 24 possible negative instances. A schema for these conditions is presented in Fig. 2.

In terms of hypotheses, there are by coincidence also 27 possible ones at the outset, such as *large black* (object), *white triangle,* etc. [In terms of the

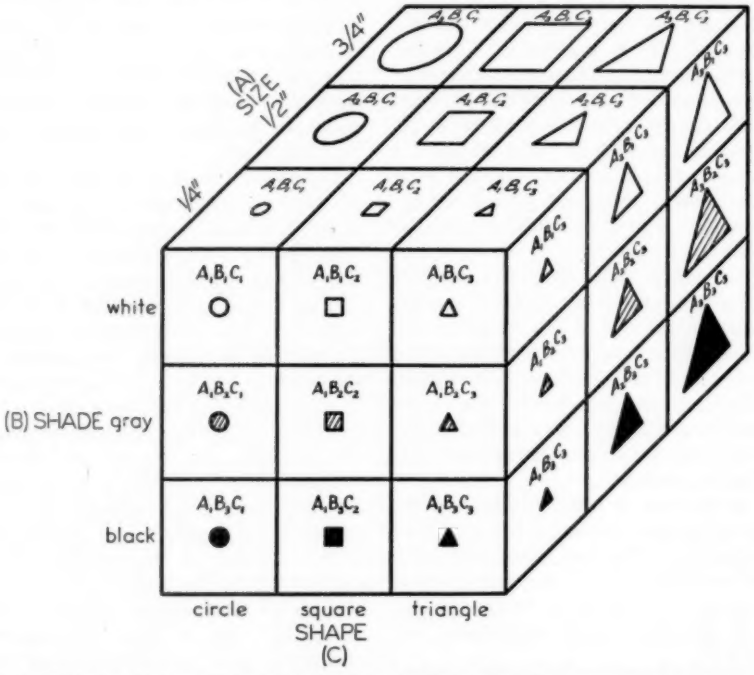

FIG. 2. Diagram showing total possible instances for three dimensions (size, shade, and shape) with *three* values of each.

diagram, these represent the sets of three blocks in a line along one of the three primary axes. There are nine such lines along each axis.] Three are consonant with the first positive instance but there is only one consonant with both the first and second positive instance, so only *two* of the three possible positive instances need be transmitted. With negative instances the first instance eliminates three hypotheses, the second, three, the third, three, etc. After seven instances there is redundancy, so that after 21 hypotheses have been eliminated the remaining 6 require a total of 3 negative instances to eliminate all but the correct hypothesis. Hence there are required a minimum of 10 negative instances. Thus *two* positive or *ten* negative instances will completely define the correct concept. [In terms of our geometrical figure two blocks will define the dimension along which adjacent blocks are lined up, but it will take 10 placements in a three-dimensional "Tit-tat-toe" to plug all but one line of blocks.]

In the foregoing example more negative instances than positive were required. But this is not always the case. Consider the situation where there are three possible dimensions and five values along each, and where the correct concept involves four values of one of the dimensions. Let us say that color, size, and shape are again involved, and that each can have any one of five values. Thus color can be red, green, yellow, blue or black, size can be very large, large, medium, small, or very small and shape can be a circle, square, triangle, star, or diamond. If one chooses as the correct concept "a colored object" [an object of *red, green, yellow,* or *blue* (but not *black*)] (of any shape or size) how many positive and negative instances will be required to transmit the correct concept to S? Analysis of this instance reveals that there are 125 possible instances, of which 100 are correct instances (positives) and 25 incorrect (negatives). S would initially have to consider 15 possible hypotheses as to the combinations of dimensions and values. *Five* positive instances would serve to specify completely the correct concept. But it would take only *two* negative instances to do the same task. It will also be noted that in either case only a small fraction of the total instances would have to be used to define the concept completely.

Now to generalize the results. It is possible to calculate from formula the number of hypotheses which can be formed and the number of positive and negative instances required to eliminate all alternative possibilities except the correct one. The model used is for concepts which employ from 2 to N dimensions with 2 to N values for each dimension. The number of relevant dimensions can vary from one to one less than the total number and the number of correct values for each relevant dimension can vary from one to one less than the total possible values. Let

$A, B, C, \ldots$ = the separate dimensions,
D = total number of dimensions to be considered,
d = total number of *relevant* dimensions,
V_i = number of possible values for dimension i ($i = A, B, C, \ldots$),
v_i = number of *correct* values for dimension i.

The total number of possible instances (I) is the number of values for the first dimension times the number for the second, third, etc.

$$I = V_A V_B V_C \cdot \cdot \cdot V_D. \qquad (1)$$

When the number of values is the same for each dimension this formula becomes

$$I = V^D. \qquad (1a)$$

The number of positive instances (I_p) is the product of the numbers of values which are correct for each dimension:

$$I_p = v_A\, v_B\, v_C \cdot \cdot \cdot v_D. \qquad (2)$$

The number of values correct is less than the total number of values for each relevant dimension but is the entire number of values for the irrelevant dimensions. So when the number of values correct is the same for each relevant dimension, the formula reduces to the product of the number of values for the relevant dimensions times the total number for the remaining dimensions:

$$I_p = v^d\, V^{D-d}. \qquad (2a)$$

The total number of negative instances (I_n) can easily be obtained by subtraction:

$$I_n = I - I_p. \qquad (3)$$

While formulae can be presented for the general case with any number of values for each dimension, the remaining formulae will be for the frequent special case where the number of values is the same for each dimension. Under these circumstances the number of possibilities, or hypotheses, which must be considered at the outset (when the number of positive or negative instances = 0) can be determined. When one specifies the number of possible dimensions and the number which are relevant, there are ${}_DC_d$ possible ways in which these can be selected. Within each relevant dimension the correct values can be selected in ${}_VC_v$ ways, and there are d such dimensions. The total number of hypotheses at the outset (H_o) is accordingly

$$H_o = ({}_DC_d)\ ({}_VC_v)^d. \qquad (4)$$

The number of hypotheses which must be considered after j number of positive instances ($H_{p=j}$) is

$$H_{p=j} = ({}_DC_d)\ ({}_{V-j}C_{v-j})^d. \qquad (5)$$

Only one hypothesis (the correct one) remains after $v + 1$ instances have been presented:

$$H_{p=v+1} = 1. \qquad (6)$$

Thus the minimum number of positive instances required (M_p) is one more than the number of correct values for the relevant dimensions.

$$M_p = v + 1. \qquad (7)$$

Another way of viewing this number is that it is the minimum necessary to display all the non-relevant dimensions at a sufficient number of values to rule out the possibility that they are relevant. With the same number of instances one is able simultaneously to specify all the correct values for the relevant dimensions. So if there are three values correct for dimensions A and B, and C is the irrelevant dimension, one must give instances with each of the three values of A and B plus a fourth instance involving the fourth value of C. The latter information indicates that C is an irrelevant dimension.

Since the initial number of hypotheses is of course identical whether one is to employ positive or negative instances ($H_{p=o} = H_{n=o}$) the number of possible hypotheses remaining after one negative instance is equal to the initial number minus the number specificed above as remaining after one positive instance:

$$H_{n=1} = H_{p=o} - H_{p=1}. \qquad (8)$$

From this point on the number of possible hypotheses does not always follow a simple pattern, because there is an indeterminate amount of overlap ("redundancy") between the hypotheses eliminated by each new instance and those which have been eliminated previously.[3] Thus the minimum number of

[3] Except for the limiting case where only one relevant dimension is involved ($d = 1$). Here $H_{n=j} = ({}_{V-j}C_v)\ ({}_DC_d)$ and $M_n = V - v + 1$.

negative instances (M_n) can be given only approximately at present. For the range of cases studied to date the number is approximately [4]

$$M_n \cong [V - (v - 1)]^d. \qquad (9)$$

A check on the formulae presented can be made by determining empirically the minimum number of positive or of negative instances and the number of possible hypotheses which remain after each successive instance has been presented. Empirical values are presented in Table 1 for a variety of illustrative cases where D's and V's are between two and five. Perfect agreement with formula is obtained for all entries except for the minimum number of negative instances (M_n), which was stated to be only an approximation. For the cases illustrated the maximum deviation between the empirical and predicted values is four instances.

Consideration of the table and formulae suggests conditions under which negative instances convey little or much information. In general, when the number of values correct remains constant, as the total number of possible values increases the minimum number of *nega-*

[4] When the number of dimensions is more than one greater than the number of values for each relevant dimension a correction term must be introduced. Robert Abelson suggests that for concepts with one value correct for each of two relevant dimensions when $D > V + 1$ the equation must be multiplied by a fraction approximately equal to $\frac{\log D}{\log (V+1)}$.

TABLE 1

ILLUSTRATIVE CASES FOR CONCEPT MODEL WHERE NUMBERS OF DIMENSIONS AND VALUES ARE SPECIFIED

Tot. No. Dimens. (D)	No. Poss. Values each Dimen. (V)	No. Relevant Dimens. (d)	No. Values for Each Relev. Dimens. (v)	Total Instances (I)	Total Positive Instances (I_p)	Total Negative Instances (I_n)	Hypotheses at Outset (H_0)	No. Hypoth. Remaining after Successive Presentation of Positive Instances ($H_{p=1, 2, 3 \ldots}$)	Min. No. Pos. Inst. Req. (M_p)	No. Hypoth. Remaining after Successive Presentations of Negative Instances ($H_{n=1, 2, 3 \ldots}$)	Min. No. Neg. Inst. Req. (M_n)
2	3	1	1	9	3	6	6	2-1	2	4-2-1	3
2	4	1	2	16	8	8	12	6-2-1	3	6-2-1	3
2	5	1	4	25	20	5	10	8-6-4-2-1	5	2-1	2
3	2	2	1	8	2	6	12	3-1	2	9-6-3-2-1	5
3	4	1	1	64	16	48	12	3-1	2	9-6-3-1	4
3	4	2	3	64	36	28	48	27-12-3-1	4	21-6-2-1	4
3	5	1	4	125	100	25	15	12-9-6-3-1	5	3-1	2
3	5	2	2	125	20	105	300	48-3-1	3	252-207-165-126-90-72-55-40-30-21-13-7-4-2-1	15
4	2	3	1	16	2	14	32	4-1	2	28-24-20-16-12-8-4-3-2-1	10
4	4	2	3	256	144	112	96	54-24-6-1	4	42-12-3-1	4
4	4	3	2	256	16	240	864	108-4-1	3	756-652-552-456-388-324-267-221-184-154-126-100-78-59-41-27-20-15-10-7-4-2-1	23
4	5	2	1	625	25	600	150	6-1	2	144-138-132-126...6-3-1	26
4	5	2	4	625	400	225	150	96-54-24-6-1	5	54-12-3-1	4
4	5	3	1	625	5	620	500	4-1	2	496-492-488-484...4-3-2-1	127
4	5	3	4	625	320	305	500	256-108-32-4-1	5	244-96-50-15-8-3-2-1	8
5	2	3	1	32	4	28	80	10-1	2	70-60-50-40-32-24-17-10-7-4-2-1	12
5	4	2	2	1024	256	768	360	90-10-1	3	270-190-120-60-42-28-18-8-2-1	10
5	4	3	2	1024	128	896	2160	270-10-1	3	1890-1630-1380-1140-1110-883-731-588-506-417-345-293-231-192-160-130-87-71-56-44-30-20-14-8-5-2-1.	27
5	5	1	1	3125	625	2500	25	5-1	2	20-15-10-5-1	5
5	5	1	4	3125	2500	625	25	20-15-10-5-1	5	5-1	2
5	5	2	4	3125	2000	1125	250	160-90-40-10-1	5	90-20-4-1	4
5	5	4	1	3125	5	3120	3125	5-1	2	3120-3115-3110-3105...5-4-3-2-1	628

tive instances required to convey the concept increases. Thus in the case where the total number of dimensions is three, and there are two relevant dimensions with a single value of each, as one increases the number of values the minimum number of positive instances stays constant at 2 but the minimum number of negative instances increases from 5 when only two values are involved to 26 when five are used. It can also be seen that as the number of relevant dimensions is increased the minimum number of negative instances increases, since the number of relevant dimensions is used as an exponent in the formula. For example, with four as the total number of dimensions and five as the number of values as one goes from one relevant dimension to two and then to three the minimum number of positive instances stays constant at 2 but the number of negatives goes from 5 to 30 and then to 127.

The conditions where the most favorable ratios of negative instance to positive exist are those in which nearly all of the values of each relevant dimension are correct. One such instance was already cited where fewer negative than positive instances were required (above, p. 467). Under these conditions one is really eliminating the values which are not involved and the negative instances do this operation more effectively than positive instances.

Combinations of Positive and Negative Instances. While the analysis of all positive or all negative series provides the greatest amount of knowledge about the information-transmitting characteristics of the two types of instance, there are many situations where both are used in the same concept-formation series. A sample case might be one where four dimensions and two values are employed, with two relevant dimensions and one value of each correct. Here two positives or five negatives are needed if all instances in either series are of the same type. However, two negatives can be substituted for one of the positives, so that a series containing one positive and two negative instances will provide the entire amount of information. Thus with size (large or small), shape (circle or square), color (red or black), and quantity (one or two of each figure) as dimensions, the concept *large circle* (of either quantity and color) can be communicated by showing a single large red circle as a positive instance, and a single large red square as one negative instance and a single small red circle as the other.

From the standpoint of the information transmitted the sequence of positive and negative instances is immaterial, but psychologically there may be some interesting problems as to the order in which the instances are transmitted. When, for example, two positive and two negative instances will define a concept, it will be instructive to see if the concept is learned equally rapidly when the positives are first (since they reduce rapidly the number of hypotheses which are possible) and then the negatives (which pin-point the possibilities) as compared with the reverse order, where a large number of hypotheses are left after the first two negative instances. An experimental test of this hypothesis will be found in Hovland and Weiss (2).

Other Models. The models which we have been discussing in this paper are of course only a small number out of the possible ones. A variety of challenging problems are suggested with other models. One simple variant is one in which the values of each dimension are interrelated in some systematic fashion. One pattern involves contiguous values with the restriction that only adjacent values are to be selected when more than one value is correct for each dimension. E.g., if two values of the size factor are involved they must be adjacent, like

small and medium, or medium and large, but not small and large. The effect of this requirement is to reduce the number of hypotheses and to require fewer instances.

Another type of model is involved in contingent relationships, typified by a situation where values 1 and 2 of dimensions A and B are correct when dimension C has value 1 but values 3 and 4 are correct when C has value 2. There are certainly some real-life concepts of this type. With some modification they may be synthesized from the model used in the present paper, but may be more efficiently handled as a separate pattern.

Variations in Procedure. Variations in the amount of prior knowledge of *S* about the number of dimensions and values to be used deserve careful exploration. Preliminary work suggests that when there is no specification concerning the dimensions to be used, negative instances are of great significance in eliminating false hypotheses concerning relevant dimensions and preventing overgeneralizing from all positive instances.

Another interesting variation in procedure involves changing the character of the information conveyed. In the examples given in the present paper and in Smoke's studies *S*s only know that a negative instance is incorrect in some respect, but the locus of the variation is not specified. Analyses could be made of the concept-learning problem where *S*s are told that the negative instance is wrong in one particular respect, or in two, etc. This may be the type of design which is sometimes involved in the use of negative instances in classroom teaching where the exact nature of the deviation from the correct concept is specified. The opposite direction of variation is one in which the values and dimensions are not invariably associated in the concept, and *S*s are only informed that the instance is "probably correct" or "probably incorrect." This type of procedure simulates many concept-learning situations in everyday life where the characterization of correctness can only be made in probability terms. Studies of this type may help bridge the gap between laboratory investigations and the complex conditions which obtain naturalistically.

Finally, in the present paper the analysis has been restricted to the ***minimum*** number of appropriately *selected* instances which are necessary to convey a particular concept. Another type of analysis would involve determining the *average* number of randomly selected instances required. Under these conditions high redundancy would be involved. It is likely that the type of set-up used by Smoke combines the uncertainties of this variation with the uncertainty produced by *S*'s not knowing the nature of the concept model. Smoke appears to have presented random-like variation of instances and the nature of the concept was not specified. Under these circumstances it would be no wonder that negative instances were found to be of so little value.

SUMMARY

Experimental studies by Smoke (**5, 6**) and others on the effectiveness of positive and negative instances in concept formation leave unanswered the extent to which the inefficiency of negative instances is due to *S*'s difficulty in assimilating information presented in this form (in terms of what the concept "is not") and the extent to which it is due to the low efficiency of negative instances in transmitting the necessary information as to the characteristics of the concept. Further experimental work will have to hold the amount of information conveyed by the two types of instances constant in order to determine the relative difficulty of learning from the two types of instance.

As preliminary to such investigations the present paper makes a theoretical analysis of one important concept model. On the basis of this analysis it is possible to determine precisely the minimum number of each type of instance required to communicate correctly the characteristics of a concept to a *S* who knows the nature of the concept model and the number and types of dimensions which are to be considered. It is shown that for the model employed the relative number of instances of the two types required varies enormously as a function of the total number of dimensions involved, the number of these which are relevant to the concept, the total number of values used for each dimension, and the number of correct values for the relevant dimensions. Examples are given of conditions where only *two* positive instances but *625* negative instances are required to define the concept, and others where *five* positive instances but only *two* negatives are required. Formulae derived theoretically are presented, together with a table giving empirical data on the minimum number of positive and negative instances required under various conditions.

In the light of the considerations presented, it is clear that the question asked in earlier concept-formation studies as to the relative effectiveness of positive and negative instances cannot be given a generalized categorical answer. Separate analyses must be made of (1) the amount of information conveyed by each type of instance under specified conditions and (2) the process of assimilating information from the two types of instance when the amount of information transmitted is equated.

REFERENCES

1. BUSS, A. H. A study of concept formation as a function of reinforcement and stimulus generalization. *J. exp. Psychol.*, 1950, **40**, 494–503.
2. HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances (in preparation).
3. OSEAS, L., & UNDERWOOD, B. J. Studies of distributed practice: V. Learning and retention of concepts. *J. exp. Psychol.*, 1952, **43**, 143–148.
4. SHANNON, C. E., & WEAVER, W. *The mathematical theory of communication.* Urbana: University of Illinois Press, 1949.
5. SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**, No. 4.
6. ——. Negative instances in concept learning. *J. exp. Psychol.*, 1933, **16**, 583–588.

[MS. received for early publication August 8, 1952]

THE BRITISH JOURNAL OF PSYCHOLOGY

(General Section)

Edited by D. W. HARDING

Vol. XLIII. Part 2 May 1952 12s. 6d. net.

P. L. BROADHURST, C. P. STONE and D. H. LAWRENCE. The effects of epanutin and electroconvulsive shock on the maze performance of rats.

B. SEMEONOFF and A. J. LAIRD. The Vigotsky Test as a measure of intelligence.

S. CROWN. The word connection list as a diagnostic test: norms and validation.

I. M. L. HUNTER. An experimental investigation of the absolute and relative theories of transposition behaviour in children.

E. BARSCHAK. Happiness and unhappiness in the childhood and adolescence of a group of women students: a comparative study between English and American girls.

M. B. SHAPIRO. Some correlates of opinions on the bringing up of children.

PUBLICATIONS RECENTLY RECEIVED.

Vol. XLIII. Part 3 August 1952 12s. 6d. net.

LEWIS F. RICHARDSON. Dr. S. J. F. Philpott's wave-theory.

S. J. F. PHILPOTT. Reply to Dr. Richardson.

DURGANAND SINHA. Behaviour in a catastrophic situation: a psychological study of reports and rumours during Darjeeling landslides.

J. P. LEES. The social mobility of a group of eldest-born and intermediate adult males.

E. C. POULTON. Perceptual anticipation in tracking with two-pointer and one-pointer displays.

PUBLICATIONS RECENTLY RECEIVED.

The subscription price per volume, payable in advance, is $6.50

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS

American Branch, 32 East 57th Street, New York 22, N.Y.